Análise do movimento de clientes em loja com base em sistemas de posicionamento

Susana Daniela Viana Rolo

Dissertação de Mestrado

Orientador na FEUP: Prof. Vera Miguéis

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Mestrado Integrado em Engenharia Industrial e Gestão

2015-07-27

Resumo

Ao longo dos últimos anos, a elevada competitividade existente no mercado obrigou os retalhistas a melhorarem os seus processos operacionais para que se conseguissem adaptar e sobreviver no meio. Uma vez homogeneizadas as operações, chega a altura das empresas se diferenciarem nos detalhes.

Apesar de serem vários os estudos que destacam a influência de um *layout* atrativo numa compra, são ainda escassas as pesquisas acerca de como medir essa eficácia.

Uma das formas de investigar o quão bem conseguido está um espaço, passa por analisar o movimento dos clientes quando visitam a loja. Para uma correta avaliação, é fundamental que se consiga perceber quais são os trajetos típicos dos consumidores ou quais as zonas quentes e frias da loja.

Este projeto surge como forma de dar resposta a essa necessidade. Foi desenvolvida uma ferramenta de análise capaz de obter os percursos efetuados, num determinado espaço comercial, partindo da potência do sinal emitida por dispositivos móveis, através de *WiFi* ou *Bluetooth*, pertencentes a clientes. O modelo implementado aplica, depois, essa informação no cálculo de métricas capazes de descreverem a eficiência do *layout.*

No processo criado, o posicionamento poderá ser feito através de dois métodos, proximidade e análise de cenário. Relativamente à análise dos trajetos, para além de se avaliar o número de pessoas que vai a uma determinada zona loja ou o tempo médio de permanência em cada zona, é ainda implementada uma funcionalidade que permite averiguar as transições entre áreas. Para esse fim, são empregues cadeias de Markov.

Todo o sistema foi criado para lidar com as diferentes características de cada espaço e, portanto, pode ser aplicado a qualquer tipo de superfície comercial. Para testar a automatização do processo, foram analisadas amostras de dados referentes a dois tipos de loja diferentes.

Client's movements analysis in store based on positioning systems

Abstract

Over the past few years, the high competitiveness of the market has forced retailers to improve their operational processes in order to adapt and survive. Since most companies already perform at high operational levels, it has come the time for these companies to differentiate themselves in the details.

Although there are several studies that highlight the influence of an attractive layout in a purchase, there are still insufficient researches on how to measure its effectiveness.

One way to check how well an area is structured is by analyzing the customers' movements when they visit the store. For a correct evaluation it is essential to be able to understand which are the typical consumers' paths or where are the hot and cold spot zones located in the store.

This project main purpose is to address that need. Therefore, it has been developed an analysis tool which is able to study the paths of customers, in a given store, based on the signal strength emitted by their mobile devices, through WiFi or Bluetooth.

The established framework applies this information to calculate metrics that will allow the definition of the layout's efficiency.

Positioning can be obtained by two methods, proximity and scene analysis. Regarding the paths' analysis, besides evaluating the number of people who go to a specific store zone or the average time each costumer spends in each zone, this application has also a feature implemented that allows the determination of the probability of transitions between areas. For that purpose, Markov chains are employed.

The whole system is designed to deal with different characteristics of each space and thus can be applied to any type of commercial surface. In order to test the process automatization, data samples of two different types of store were analyzed.

Agradecimentos

Ao orientador da InovRetail, Eng.º João Guichard, por todo o apoio e motivação transmitida no decorrer deste projeto.

À orientadora da FEUP, Prof. Vera Miguéis, pela disponibilidade e acompanhamento prestado ao longo desta etapa.

À minha família, por estar presente em todos os momentos da minha vida, em especial aos meus pais, por todas as oportunidades que me proporcionaram.

Aos meus amigos, pela paciência e confiança transmitida durante esta fase e por todas as experiências vivenciadas durante todo o meu percurso académico.

Índice de Conteúdos

Índice de Figuras

Índice de Tabelas

1 Introdução

A presente dissertação foi realizada no âmbito do Mestrado Integrado em Engenharia Industrial e Gestão da Faculdade de Engenharia da Universidade do Porto, na InovRetail. Neste capítulo irá ser feita uma breve apresentação da empresa onde foi realizado o projeto, bem como do projeto em si. Abordar-se-á ainda a metodologia seguida para a sua realização e a estrutura do documento.

1.1 Apresentação da InovRetail

A InovRetail é uma empresa portuguesa que se dedica à Investigação & Desenvolvimento de soluções de base tecnológica e que tem como missão melhorar a experiência de compra dos clientes, por intermédio de ambientes de loja mais apelativos, dinâmicos e eficientes, com retorno mensurável para os retalhistas.

Está no mercado desde 2011, tem a sua sede na UPTEC – Parque da Ciência e Tecnologia da Universidade do Porto e pretende tornar-se uma referência internacional no desenvolvimento de soluções inovadoras e de elevado retorno para o retalho. Tem como valores a qualidade, a inovação, a capacidade de entrega e a aposta na internacionalização.

1.2 A importância do projeto na InovRetail

O objetivo da InovRetail é ajudar os retalhistas a melhorar os seus espaços de maneira a que se consiga influenciar de forma positiva o comportamento do consumidor em loja.

Para se compreender o estado atual e o impacto de possíveis medidas, torna-se essencial conhecer os trajetos feitos pelos consumidores num e noutro momento. Só desta forma se poderá perceber que alterações foram bem conseguidas e que aspetos necessitam de aperfeiçoamento.

Nos últimos anos foram vários os projetos desenvolvidos pela empresa com esse fim. Apesar do objetivo ser o mesmo, as lojas em análise diferiam umas das outras, tanto no *layout* como no tipo de produtos vendidos. Cada caso era estudado de forma individual e cada análise era realizada tendo em conta as características de cada estabelecimento comercial.

Como cada modelo criado é feito de forma aplicada diretamente à loja em questão, não há a possibilidade de o reutilizar. Em cada novo projeto, é sempre necessário gastar tempo e recursos no desenvolvimento de uma nova ferramenta de análise que tenha em conta as restrições daquele espaço. Existe, por isso, a necessidade de automatizar este processo de forma a torná-lo mais eficiente.

1.3 Objetivos do projeto

De forma sintética, a realização deste projeto visa conceber um modelo analítico capaz de avaliar o movimento dos consumidores em qualquer tipo de espaço comercial.

A ferramenta criada deverá contemplar as duas fases do processo. Primeiramente, é necessário que se convertam os sinais emitidos pelos dispositivos móveis dos clientes, via *Wi-Fi* ou *Bluetooth*, em posições e, só depois, é que essas posições deverão ser analisadas e traduzidas em indicadores que permitam obter informações relevantes para o negócio em questão.

É esperado que o posicionamento possa ser feito através de dois métodos: proximidade e análise de cenário e que englobe todas as etapas necessárias para uma correta filtragem dos dados.

Outro dos desafios do projeto será a criação de novas métricas de análise das posições. Assim, o modelo deverá ser capaz de calcular não só os indicadores já estimados pela empresa anteriormente, como o tempo médio de permanência numa zona ou a percentagem de pessoas que passa por determinado sítio, mas também novas métricas que elevem o grau de conhecimento extraído dos dados relativos ao movimento dos clientes em loja.

Por fim, e para que a solução implementada continue a ser usada no futuro, é necessário que a ferramenta desenvolvida seja fácil de entender e de utilizar. Para que não surjam dúvidas no seu funcionamento, deverá ser criado também um manual de utilização.

1.4 Método seguido no projeto

No início do projeto foi definido um macro plano a ser seguido na resolução do problema proposto. O planeamento decidido é apresentado na Figura 1.

Nº	Atividade	Início	Fim	Duração	fev/15	mar/15	abr/15	mai/15	jun/15
1	Compreensão do Problema	09/02/2015	27/02/2915	15d					
2	Desenvolvimento do Modelo de Posicionamento	23/02/2015	27/03/2015	25d					
3	Desenvolvimento do Modelo de Análise de Trajetos	27/03/2015	29/05/2015	46d					
4	Otimização dos Modelos	29/05/2015	12/06/2015	11d					

Figura 1 - Diagrama de *Gantt* com as fases mais importantes do projeto.

Primeiramente, foi feita uma análise dos processos já realizados na empresa, tanto a nível de posicionamento como de métricas já calculadas. Foi possível compreender com mais detalhe o problema apresentado e foi feito um levantamento bibliográfico acerca do tema.

Para o desenvolvimento do modelo de posicionamento, foi necessário fazer uma pesquisa mais aprofundada acerca dos sistemas de posicionamento existentes. Depois de obtido esse conhecimento, foi desenvolvida a ferramenta de análise em questão.

A etapa seguinte envolveu a análise dos trajetos e teve como objetivo a criação de um modelo capaz de o fazer. Esta foi a fase mais longa do projeto e pretendeu-se que fossem calculadas métricas da loja relativas ao movimento dos consumidores em loja. Para se realizar este último ponto, foi também necessário fazer uma nova pesquisa bibliográfica, a fim de perceber que ferramentas poderiam ser utilizadas para extrair conhecimento deste tipo de dados, para além das já usadas na empresa.

Por fim, foi necessária uma otimização de ambos os modelos. Esta melhoria focou-se sobretudo nos tempos de análise dos dados.

1.5 Estrutura da dissertação

Depois de feita a introdução, em que se apresentou a empresa envolvida e se caraterizou o projeto e os seus objetivos, o capítulo seguinte contempla uma revisão teórica relativa aos sistemas de posicionamento em ambientes interiores existentes e aos métodos de análise dos dados utilizados. No terceiro capítulo, é apresentado o problema em estudo de forma detalhada e no capítulo quatro é caracterizada a solução implementada. A descrição dos dados amostrais e a sua análise é feita no capítulo cinco. Por fim, no sexto capítulo, são descritas as principais conclusões inferidas ao longo do desenvolvimento deste projeto assim como possíveis melhorias.

2 Enquadramento teórico

Neste capítulo será apresentado um breve enquadramento teórico que visa os temas inerentes ao projeto.

2.1 Sistemas de Posicionamento

O retalho é um tipo de setor com um elevado nível de competição (Singh *et al*, 2014). Segundo Cil (2012), o sucesso das empresas deste setor é fortemente influenciado pela sua capacidade em compreender o comportamento do consumidor.

Kröckel *et al* (2011) afirma que, no passado, as vantagens competitivas eram conseguidas apenas com um bom portfólio de produtos e bons preços mas, hoje em dia, os retalhistas são forçados a arranjar novas estratégias para conseguirem cativar os clientes. Para preservar e melhorar a retenção do consumidor é necessário que estes se foquem nas preferências e aspetos que influenciam a decisão de compra (Singh *et al*, 2014)

Ao contrário das lojas *online*, que conseguem obter facilmente dados relativos ao tempo gasto no website ou à sequência de passos em cada visita, os retalhistas com lojas físicas tradicionais carecem de informação acerca dos seus clientes. Por essa razão, torna-se necessário o desenvolvimento de ferramentas que permitam esse tipo de análises (Kröckel *et al*, 2011).

Phua *et al* (2015) refere que o uso de tecnologias para rastrear o comportamento do consumidor tem aumentado nos últimos anos na indústria do retalho. Com a informação obtida, é possível perceber como se deslocam realmente os clientes na loja: se vão a todas as áreas ou se passam de uma zona para outra de forma mais direta, se seguem um padrão dominante ou se são trajetos heterogéneos (Cabanes *et al* 2009). É possível ainda verificar quais as zonas mais visitadas ou qual o tempo médio de uma visita (Phua *et al*, 2015).

O *layout/design* de uma loja é um dos fatores que afeta o comportamento de compra dos consumidores (Lewison, 1994; Kröckel *et al*, 2011; Singh *et al*, 2014; Phua *et al*, 2015). Quando bem planeado, pode encorajar os visitantes a passarem por mais áreas, vendo uma maior variedade de produtos e aumentando assim a probabilidade de compra (Levy, 2001). Desta forma, o acesso à informação relativa ao deslocamento do consumidor em loja permite não só percebê-lo como também verificar a eficiência do *layout* (Kröckel *et al*, 2011). Esta informação poderá ser usada ainda para planear a alocação de funcionários (Phua *et al*, 2015).

Ao longo do tempo, os dados relativos aos trajetos dos consumidores têm sido obtidos através de várias formas. Inicialmente, o método utilizado traduzia-se por seguir fisicamente as pessoas na loja de forma a registar os seus movimentos (Farley, 1996). A limitação deste processo e o avanço da tecnologia fez com que novas técnicas surgissem para avaliar o comportamento do consumidor.

2.1.1 Tecnologias existentes

Os sistemas de análise baseados em vídeo têm sido a abordagem mais frequente para capturar o comportamento humano em ambientes fechados (Millonig *et al*, 2008; Sorensen, 2009; Kröckel *et al*, 2011; Scamell-Katz, 2012).

O reconhecimento de objetos, baseado em vídeo, é ainda um campo ativo na investigação, mas ainda não oferece soluções fiáveis e de baixo custo para o retalho (Yan *et al*, 2008). Para além de não ser uma solução viável (Versichele *et al*, 2012), exige muito esforço computacional e não garante uma precisão elevada (Yan, 2010; Yan *et al*, 2008).

Mais recentemente, têm sido usados sistemas baseados em rádio frequência para a identificação dos percursos dos consumidores. Como usam uma transmissão eletromagnética, este tipo de sistemas tem capacidade para abranger uma grande área e para ultrapassar obstáculos opacos, tais como paredes ou pessoas (Zhang *et al*, 2010). Um dos métodos já implementados neste contexto consiste no uso de etiquetas que permitem a identificação por rádio frequência (RFID) colocadas nos carrinhos ou cestos de compras (Yan *et al*, 2008; Yan, 2010; Utsch *et al*, 2012). Para além desta tecnologia, o movimento dos consumidores em loja tem sido também analisado tendo por base as tecnologias *Bluetooth* e *Wi-Fi*. Estas tecnologias permitem registar o movimento individual do consumidor a um baixo custo, ao contrário dos sistemas baseados em vídeo ou dos que utilizam a identificação por rádio frequência (Delafontaine *et al*, 2012). Além disso, permitem a localização de qualquer dispositivo compatível, sem ser para isso necessário instalar *software* extra ou manipular o *hardware* (Farid *et al*, 2013). Para além do seu baixo custo, este método apresenta outras vantagens como é o caso de se poder recolher uma grande quantidade de dados num determinado período de tempo e o facto da informação recolhida ser de natureza não evasiva e não identificada, não dependendo da cooperação ou memória do comprador (Phua *et al*, 2015).

O endereço MAC de cada dispositivo atua como um identificador único do aparelho na rede e pode ser usado para ligar diferentes registos do mesmo utilizador, de forma a gerar a sua trajetória. Como os indivíduos rastreados não são abordados, os dados não contêm informação sociodemográfica, o que lhes permite permanecerem anónimos (Versichele *et al*, 2014; Phua *et al*, 2015).

Apesar das vantagens, este tipo de sistemas possui também algumas limitações, como por exemplo a atenuação do sinal devido a obstáculos presentes na área e o elevado consumo de energia nos dispositivos móveis (Farid *et al*, 2013). Os trajetos obtidos por *Bluetooth* ou *Wi-Fi* podem também ser incompletos ou inconsistentes devido à falha de dados nos recetores (obstrução de sinal, perda de sinal) e a problemas no aparelho (vida das baterias limitada, *Bluetooth/Wi-Fi* desligado pelo utilizador) (Delafontaine *et al*, 2012).

Alguns autores mencionam o facto dos métodos de rastreamento que usam este tipo de tecnologias poderem produzir amostras de dados limitadas e possivelmente enviesadas, uma vez que não só é necessário que o consumidor possua um aparelho com *Bluetooth/ Wi-Fi* como também é requisito que o tenha ligado para que seja detetado (Versichele *et al*, 2012). Estas limitações podem levar a que certos segmentos da população possam estar sub- ou sobre-representados (Rice *et al*, 2003).

Na comparação entre *Bluetooth* e *WiFi*, verifica-se que são mais as pessoas que têm um aparelho habilitado para *Wi-Fi* (Abedi *et al*, 2013; Phua *et al*, 2015). Além disso, com a crescente disponibilidade e aceitação do público de *spots Wi-Fi*, verificou-se também que é mais provável que os consumidores tenham *Wi-Fi* ligado do que o *Bluetooth* (Phua *et al*, 2015).

2.1.2 Posicionamento

O desafio na obtenção de posição em ambientes fechados através de radiações eletromagnéticas reside no facto de esta ser altamente influenciada por erros de múltiplos caminhos *(multipath)*, pela baixa probabilidade de existência de uma linha de visão entre o emissor e recetor (*Non-Light-of-Sight)*, pela presença de pessoas que se deslocam e que modificam o canal de propagação do sinal e pela densidade de obstáculos que causam um elevada atenuação e espalhamento do sinal (Sarkar *et al*, 2003; Liu *et al*, 2007; Mainetti *et al*, 2014).

Para se conseguir obter uma posição através da tecnologia *Bluetooth/Wi-Fi* é necessária a instalação de equipamentos capazes de registar o valor da potência do sinal (*Received Signal Strenght Intencity, RSSI)* emitido na área. Sempre que um aparelho é detetado, é registado o seu endereço MAC, a hora de deteção e o valor de RSSI captado pela antena (Delafontaine *et al*, 2012). Como se pode verificar na Figura 2, o valor de RSSI detetado apresenta uma correlação negativa com a distância entre a antena e o aparelho detetado (Farid *et al*, 2013).

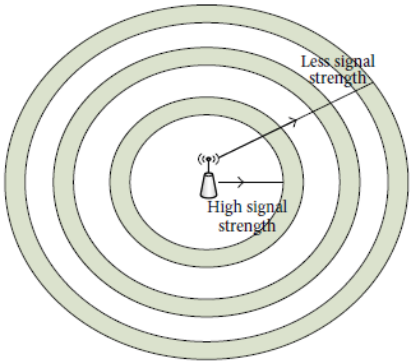


Figura 2 - Padrão da Potência do Sinal (Farid *et al*, 2013).

Depois de lido o sinal é necessário convertê-lo em posição. Esta conversão com base na potência do sinal recebido poderá ser feita por triangulação, proximidade ou análise de cenário (Liu *et al*, 2007).

A triangulação consiste no uso das propriedades geométricas dos triângulos para determinar a posição alvo (Farid *et al*, 2013). É necessário que o dispositivo seja registado por três antenas naquele momento para que se possa determinar a sua posição. A posição é obtida cruzando a distância calculada entre o emissor com cada um dos recetores tendo em conta a atenuação do sinal (Lopes, 2014). Quando comparada com o método de análise de cenário, esta técnica não apresenta um maior grau de precisão, uma vez que é afetada por erros de múltiplos caminhos e pela posição das antenas (Liu *et al*, 2007; Farid *et al*, 2013).

Os algoritmos de proximidade fornecem uma informação da localização simbólica. Normalmente é necessário uma grande densidade de antenas, sendo que cada uma tem uma localização conhecida. Quando o aparelho é detetado por uma única antena, é-lhe atribuído a localização dessa antena. Quando é detetado por mais do que uma antena, a sua posição será a localização da antena com uma potência de sinal mais elevada. É um método relativamente simples de implementar (Liu *et al*, 2007; Farid *et al*, 2013). Este método tem sido um dos mais usados no rastreamento de pessoas devido à sua simplicidade e ao facto de não necessitar de outro tipo de dados para além dos recolhidos pelas antenas (Delafontaine *et al*, 2012; Versichele *et al*, 2012; Ellersiek *et al*, 2013; Versichele *et al*, 2014)

Por fim, os métodos de análise de cenário traduzem-se por algoritmos que estimam a localização do dispositivo comparando a potência do sinal registado pelas antenas naquele momento com valores previamente mapeados. Assim, este método é composto por duas fases, *offline* e *online*. Na fase *offline*, é feito um levantamento local do ambiente a ser avaliado. As coordenadas e as respetivas potências do sinal de cada localização, para cada antena, são recolhidas. Durante a fase *online*, a técnica de posicionamento local usa o conjunto de sinais observados e compara-os com os valores recolhidos, previamente, de forma a estimar a melhor posição (Liu *et al*, 2007; Farid *et al*, 2013; Lopes, 2014). O vetor relativo à potência do sinal de um dispositivo num determinado período de tempo (fase online) é representado

por (RSS_1, RSS_2, …, RSS_n) onde RSS é a potência do sinal recebido de cada antena e *n* corresponde ao número total de antenas. Para cada posição (*x*,*y*) possível é determinada a distância euclidiana através da expressão 2.1.

$$Dxy = \sum_{i=1}^{n} \sqrt{(RSSi - RSSxy)^2} \quad (2.1)$$

Onde:

Dxy, é a distância euclidiana
RSSi, é o valor de potência lido pela antena i, na fase *online*
RSSxy, é o valor de potência registado na posição *xy* pela antena i, na fase *offline*
n, é o número total de antenas

A posição atribuída ao aparelho será aquela que apresente uma menor distância euclidiana (Liu *et al*, 2007; Bouet *et al*, 2008; Khodayari *et al*, 2010; Silva *et al*, 2011)

Apesar da precisão da localização ser influenciada pelo facto de a força do sinal poder ser afetada pela difração, reflexão e espalhamento na propagação em ambientes interiores (Farid *et al*, 2013), esta é uma das técnicas mais utilizadas em sistemas de posicionamento interior uma vez que é a que apresenta melhor relação custo/precisão (Kaemarungsi e Krishnamurthy, 2004; Beder *et al*, 2011).

2.2 Processo de Descoberta de Conhecimento

Extraída a informação relativa ao posicionamento dos indivíduos em ambientes interiores, é fulcral uma análise cuidada desta informação de forma a extrair conhecimento que possa servir de apoio à tomada de decisão.

O processo de extração de conhecimento de um grande volume de dados é designado por *Knowledge Discovery in Databases (KDD)* e é constituído por várias etapas como pode ser observado na Figura 3. As fases são: seleção dos dados; pré-processamento e limpeza dos dados; transformação dos dados; Mineração de Dados (*Data Mining*); e interpretação e avaliação dos resultados (Fayyad *et al*, 1996).

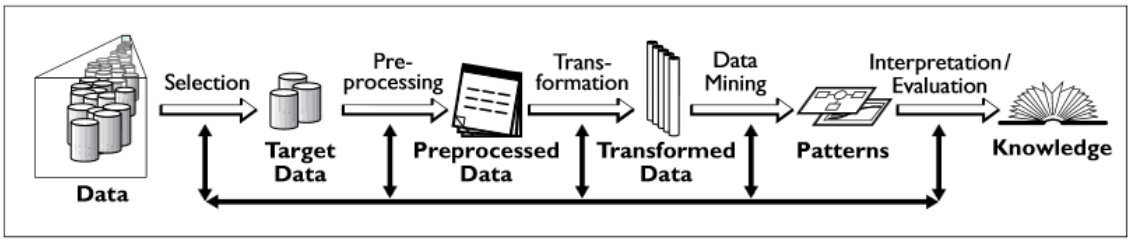


Figura 3 - Etapas do processo de descoberta de conhecimento (Fayyad *et al*, 1996).

De seguida será feita uma descrição das várias atividades presentes no processo (Raju *et al*, 2014):

Seleção de dados: Esta fase inclui o estudo do domínio da aplicação e a seleção dos dados. O objetivo é contextualizar o projeto nas operações da empresa, entendendo a linguagem do negócio e definindo metas. Nesta fase é necessário avaliar os dados a serem selecionados, os atributos relevantes e o período de tempo a ser considerado.

Pré-processamento dos dados: Nesta fase são incluídas operações básicas, tais como a remoção de *outliers*, a recolha da informação necessária para modelar o processo de filtragem ou a definição da estratégia a seguir para lidar com a falta de atributos nos dados. Nesta fase é

também definido qual será o sistema de gestão da base de dados, o tipo de dados considerados, e o esquema e mapeamento de valores em falta ou desconhecidos.

Transformação dos dados: Esta fase consiste no processamento dos dados com o objetivo de os converter nos formatos apropriados para que depois se possam aplicar algoritmos de *data mining*. As transformações mais comuns são: normalização dos dados, agregação.

Alguns algoritmos apenas lidam com dados quantitativos ou qualitativos e, por isso, poderá ser necessário transformar dados qualitativos em quantitativos e vice-versa.

Mineração de Dados: Nesta fase o objetivo é descobrir padrões nos dados já preparados. Vários algoritmos são avaliados com o objetivo de identificar qual o mais apropriado para cada tarefa específica. O algoritmo selecionado será aplicado aos dados globais para que se descubram relações indiretas ou outro tipo de padrões.

Interpretação/Avaliação: O objetivo desta fase é interpretar os padrões descobertos e avaliar a sua utilidade no âmbito do negócio. Poderá ser descoberto que alguns atributos relevantes tenham sido ignorados na análise e, por isso, o processo terá que ser refeito com um conjunto de atributos atualizado.

2.2.1 Mineração de Dados

A mineração de dados é um dos passos fundamentais no processo de descoberta de conhecimento e consiste em explorar e analisar enormes bases de dados com o objetivo de identificar padrões e regras importantes para resolver um problema. Associação, classificação, segmentação, previsão, regressão, sequenciamento, descoberta e visualização são as principais técnicas utilizadas. Neste projeto a segmentação será uma ferramenta particularmente importante.

2.2.1.1 Segmentação

A segmentação ou *clustering* é uma operação não-supervisionada. Neste tipo de operações os algoritmos criam automaticamente os seus grupos de classificação. Esta técnica deve ser usada quando se pretende encontrar grupos de dados que partilhem da mesma forma certos atributos. Após a utilização desta técnica, os resultados poderão ser usados para sintetizar o conteúdo de cada segmento da base de dados considerando apenas as características mais comuns de cada grupo. O algoritmo *k-means* é um dos algoritmos mais utilizados na segmentação (Oliveira, 2014).

O algoritmo *k-means* é normalmente usado para partir automaticamente um conjunto de dados em k grupos. A segmentação é feita através da minimização da soma da raiz quadrada da distância entre cada elemento e o correspondente centróide do *cluster (Teknomo, 2006)*. Este algoritmo começa por selecionar os centróides dos k clusters iniciais e, de seguida e de forma iterativa, vai redefinindo-os da seguinte forma (Jain, 2010):

1. Cada instância d_i é atribuída ao *cluster* que apresenta um centróide mais próximo.
2. O centróide de cada *cluster* C_j é depois atualizado com a nova média das instâncias que o constituem.

O algoritmo converge quando não existe mudança na atribuição de instâncias aos *clusters*.

2.3 Cadeias de Markov

Para se estudar as transições entre as diferentes posições dos indivíduos em ambientes interiores as cadeias de Markov têm imenso potencial. As cadeias de Markov traduzem um processo estocástico cujo conjunto de estados possíveis do sistema é discreto. Os Processos Estocásticos estudam a forma como variáveis aleatórias evoluem ao longo do tempo. Os processos de Markov consideram que as distribuições de probabilidades para estados futuros dependem unicamente do estado presente e, por isso, desconsideram como o processo chegou a tal estado.

Neste método são definidas probabilidades de transição para descrever a forma como o sistema em estudo evolui de um período para o seguinte. O objetivo é conhecer a probabilidade do sistema estar num determinado estado num período de tempo futuro.

As cadeias de Markov envolvem uma matriz de transição que contem informação relativa às probabilidades de transição entre dois quaisquer estados do sistema, possuindo uma linha e uma coluna para cada estado do sistema. Cada elemento da matriz (P_{ij}) denota a probabilidade do sistema transitar de um qualquer estado i num determinado período (t) para um qualquer estado j no período seguinte (t+1). Os valores de p_{ij} são obrigatoriamente não-negativos e a soma dos valores de cada linha da matriz será sempre igual a um.

A informação contida numa matriz de transição pode ser representada graficamente através de um diagrama de transição, tornando a sua interpretação mais simples.

Através das equações de Kolmogorov – Chapman (2.2) é possível calcular a probabilidade de transição do estado i para o estado j em n passos (p_{ij}^{n}). Essa probabilidade pode ser obtida, recursivamente, através da probabilidade de transição de uma etapa.

$$p_{ij}^{(n+m)} = \sum_{k=0}^{\infty} P_{i,k}^{(n)} P_{k,j}^{(m)} \text{ para todos } n,m \geq 0, \text{ e todos } i,j. \tag{2.2}$$

O estado *j* de uma cadeia de Markov pode ser classificado como acessível quando for possível transitar de *i* para *j* num número finito de passos. Designam-se por estados transitórios aqueles que, ao longo do processo de evolução do sistema, podendo ser visitados numa fase inicial, acabam mais cedo ou mais tarde, por serem definitivamente abandonados. Os estados que não são transitórios dizem-se recorrentes. Um estado pode ainda ser classificado como absorvente se, a partir do momento em que é atingido, nunca mais for abandonado.

Existem dois grandes tipos de cadeias de Markov: cadeias regulares e absorventes.

A regularidade de uma cadeia com matriz de transição *P* traduz-se no facto de existir um número finito n tal que todos os elementos de P^n sejam positivos. Para uma matriz com *N* estados, se os elementos da matriz P^{N^2-2N+2} forem todos positivos então o mesmo se passará com outras potências de *P* de ordem superior.

Nas cadeias regulares, à medida que *N* aumenta, todas as linhas da matriz P^n convergem para o mesmo vetor de linha de probabilidade. O conjunto desses valores forma o vetor de equilíbrio da cadeia de Markov, *v*. Cada componente do vetor, v_i, representa, no longo prazo, a proporção de vezes que o sistema se encontra no estado *i*. Isto porque, à medida que o número de transições aumenta, a probabilidade de ocorrência de um qualquer estado depende cada vez menos do estado inicial do sistema. Quando n tende para infinito, P^n tende para uma matriz *W* na qual cada linha é constituída por um mesmo vetor de probabilidade, *v*. Para qualquer vetor linha de propriedade verifica-se a igualdade apresentada em 2.3.

(2.3)

$$\lim_{n\to\infty} \pi . \mathrm{P}^n = v$$

Onde:

n, é o número de estados da matriz
π, é a probabilidade do sistema de encontrar inicialmente em cada um dos estados
P, é a matriz de transição
v, é o vetor de equilíbrio

Cada cadeia de Markov regular possui apenas um vetor de equilíbrio, sendo este o único que satisfaz a condição representada em 2.4.

$$v.P = v \tag{2.4}$$

Onde:

v, é o vetor de equilíbrio
P, é a matriz de transição

As cadeias de Markov podem ainda receber a designação de absorventes caso sejam compostas por estados transitórios e um ou mais estados absorventes. Nestas cadeias é possível atingir qualquer estado absorvente a partir de qualquer estado transitório.

Estudando o comportamento deste tipo de cadeias na fase transitória é possível obter o número médio de vezes que o sistema visita cada estado transitório, o número médio de passos até à absorção e a probabilidade de o sistema ser absorvido num determinado estado.

Para calcular o número médio de vezes que o sistema visita cada estado transitório é necessário reordenar os estados, começando pelos estados absorventes e passando, depois aos estados transitórios. Desta forma, é obtida a matriz de transição apresentada em 2.5, na sua forma canónica.

$$\left[\begin{array}{c|c} I & 0 \\ \hline R & Q \end{array}\right] \tag{2.5}$$

Onde:

I, é uma matriz identidade
0, é uma matriz de zeros
R e *Q*, são matrizes que dependem da cadeia particular

Para se obter a matriz fundamental, que representa o número médio de vezes que o sistema visita o estado transitório j, quando o processo se inicia no estado i, é necessário usar a expressão 2.6.

$$N = (I - Q)^{-1} \tag{2.6}$$

Onde:

N, é a matriz fundamental
I, é uma matriz identidade da mesma dimensão da matriz Q
Q, é uma matriz que depende da cadeia particular

O número médio de passos até à absorção é igual à soma dos números médios de vezes em que o sistema está nos diferentes estados transitórios.

Para além disso, aplicando a expressão 2.7, é ainda possível obter a probabilidade do sistema ser absorvido em cada um dos estados absorventes (j), partindo dum determinado estado transitório (i).

$$A = N.R = \lfloor aij \rfloor \quad (2.7)$$

Onde:

N, é a matriz fundamental
R, é uma matriz que depende da cadeia particular I
$a_{ij,}$ probabilidade do sistema ser absorvido num estado absorvente j, partindo de um estado transitório i

Na área do retalho, este método foi já aplicado por Kröckel *et al* (2011) com dados relativos aos movimentos de clientes num supermercado. Neste estudo, foram analisados os caminhos mais frequentes e obtiveram-se as zonas mais e menos visitadas do espaço, conhecimento que permite aos gestores de loja analisarem e otimizarem o *layout* da superfície comercial.

3 Descrição do problema

Com a crescente competitividade na indústria do retalho torna-se cada vez mais essencial perceber qual o comportamento dos consumidores em loja. Este tipo de análise permite não só avaliar a eficácia do *layout/design* escolhido como também obter informação extra relativa aos interesses dos clientes.

São cada vez mais as marcas que percebem a importância deste conhecimento no planeamento do espaço da loja e na sua gestão operacional e, por isso, esta é uma problemática que a InovRetail pretende solucionar.

Até ao momento, cada problema é estudado pela empresa de forma particular. Os métodos usados são específicos da loja em análise e os modelos desenvolvidos não permitem a sua reutilização em espaços com características diferentes.

Surge, assim, a necessidade de se criar uma ferramenta de análise transversal que possa ter em conta as restrições de *layout* presentes em cada tipo de estabelecimento e que, por isso, possa ser aplicada em todos os projetos.

O processo de análise dos trajetos de clientes num espaço comercial pode ser dividido em duas etapas. Numa primeira fase, o objetivo é obter as posições efetivas de cada visitante. A fase seguinte diz respeito à análise dessas posições de forma a extrair conhecimento válido e interessante para o negócio.

3.1 Posicionamento

O sistema de posicionamento usado pela InovRetail tira partido do uso difundido de dispositivos móveis com capacidade para estabelecer uma ligação *Wi-Fi* ou *Bluetooth*. Este tipo de sistemas de posicionamento por rádio frequência usa o valor da potência de sinal (*Received Signal Strenght Indicator*, RSSI) emitida por cada aparelho para obter a posição do mesmo.

Em cada zona da loja é colocada uma antena, com um identificador associado, que é responsável por comunicar com cada dispositivo disponível e registar o seu RSSI. As características dessa antena dependerão do tipo de tecnologia que se decida utilizar (*Wi-Fi* e/ou *Bluetooth*). Em qualquer dos casos, a partilha de informação acontece de forma contínua, sendo que quanto mais próxima a pessoa estiver da antena, maior será o valor lido.

Cada antena regista informação relativa à data e hora de cada leitura, ao endereço MAC do aparelho e ao valor de RSSI lido. Esta informação é depois consolidada numa base de dados da empresa.

O processo de consolidação engloba uma primeira filtragem. Endereços MAC que sejam registados durante um período a que a loja esteja encerrada não pertencem, com certeza, a dispositivos de clientes e, por isso, são imediatamente eliminados. Os Endereços MAC dos funcionários da loja são recolhidos previamente e eliminados também nesta etapa.

Além disso, as antenas instaladas na loja não captam os valores de RSSI em intervalos de tempo constantes uma vez que é frequente, neste tipo de sistemas de posicionamento, haver obstrução ou perda de sinal. Por essa razão, e para que haja uma uniformização dos dados, é feito um agrupamento das leituras. Registos de um determinado dispositivo móvel, lidos pela mesma antena, no espaço de cinco segundos, são convertidos num só através da sua média.

Na Tabela 1 e na Tabela 2, é possível observar um exemplo dos dados antes e depois de serem consolidados, respetivamente. Como se pode constatar, a hora é também editada e os dois algarismos correspondentes ao segundo tomam valores entre 00 e 11 consoante o período de cinco segundos a que pertença o valor lido. Além disso, tanto os registos da data como da hora são convertidos num único número para facilitar a sua ordenação.

Na base de dados é ainda guardado o identificador da loja e o tipo de tecnologia utilizada para receber o sinal emitido pelos dispositivos, *Wi-Fi* ou *Bluetooth.*

Tabela 1- Informação recolhida antes da sua consolidação.

Endereço MAC	Identificador da Antena	Valor RSSI	Data de Registo	Hora de Registo
1C:BA:8C:20:E2:FA	38	-82	24/10/2014	11:00:20
1C:BA:8C:20:E2:FA	38	-81	24/10/2014	11:00:24
1C:BA:8C:20:E2:FA	38	-80	24/10/2014	11:00:24
34:B1:F7:D1:40:6A	38	-60	24/10/2014	11:00:37
34:B1:F7:D1:40:6A	38	-64	24/10/2014	11:00:40

Tabela 2 - Informação consolidada.

Endereço MAC	Identificador da Antena	Valor RSSI	Data de Registo	Hora de Registo	Intervalo de Segundos
1C:BA:8C:20:E2:FA	38	-81	20141024	110004	4
34:B1:F7:D1:40:6A	38	-62	20141024	110007	7

Na fase seguinte do posicionamento, a informação guardada na base de dados é usada na reconstrução dos trajetos dos visitantes da loja. É nesta fase que existe a necessidade de uma homogeneização do processo.

Um dos problemas é o facto de não haver ainda uma automatização na leitura dos registos diretamente da base de dados. Até ao momento, esses registos têm sido primeiramente consultados e depois guardados num ficheiro. O formato desse ficheiro também não tem sido constante ao longo dos projetos já desenvolvidos. Como o número total de leituras, de cada loja, excede o limite máximo de linhas do *Excel*, é necessário que se faça mais do que uma consulta à base de dados. A pessoa responsável terá de dividir o intervalo de tempo em períodos menores, certificando-se que nenhum tem um total de registos superior ao limite máximo. É uma tarefa repetitiva e morosa que deverá ser automatizada.

Além disso, na conversão dos valores de RSSI em posição, há parâmetros que dependem do *layout* da loja ou do número e posição das antenas. De caso para caso, essas restrições têm sido aplicadas de forma estática, tornando impossível a reutilização das ferramentas de análise desenvolvidas. Para que a eficiência do processo seja melhorada, é essencial que se crie um modelo de posicionamento dinâmico e adaptável a todos os espaços.

Na obtenção da posição, a InovRetail já usou nos seus projetos os três métodos possíveis: análise de cenário, proximidade e triangulação. Há, por isso, a necessidade de se perceber

qual/quais dos métodos deverão ser implementados na solução desenvolvida para que se consiga analisar qualquer tipo de loja.

Uma das etapas essenciais no processo de posicionamento é a filtragem dos dados. O objetivo desta etapa é garantir que os trajetos analisados são de pessoas que realmente estiveram na loja. Para isso, têm sido aplicadas restrições relativas ao número mínimo e máximo de leituras por endereço MAC, ao valor de RSSI mínimo necessário para que uma leitura seja considerada e ao tempo decorrente entre a primeira e última leitura de cada trajeto. Além disso, é ainda verificado se os trajetos registam uma entrada e uma saída da loja e se houve continuidade na receção do sinal. O valor destas restrições tem dependido, mais uma vez, de caso para caso.

3.2 Análise dos Trajetos

Depois de obtidos os trajetos dos consumidores em loja, o passo seguinte passa por extrair conhecimento relevante dos mesmos. Até à data, as posições foram estudadas com o fim de obter a percentagem de pessoas que passa em cada área (retenção) e o tempo médio de permanência (penetração) numa determinada zona da loja. Além destes indicadores, é ainda estudado o tempo total de permanência dos clientes na superfície comercial.

Este tipo de estudo é feito, mais uma vez, de forma não automatizada. Como os resultados do posicionamento não são igualmente formatados, não existe uma ferramenta capaz de transformar diretamente esses dados numa análise geral do comportamento dos clientes em loja.

Para além de ser indispensável o desenvolvimento de um modelo de análise geral, capaz de processar os resultados obtidos no posicionamento, é também importante alargar essa análise para que se possa extrair mais informação a partir dos dados já disponíveis.

Um dos temas a ser desenvolvido está relacionado com a identificação de padrões nos trajetos. É essencial que se perceba quais as transições entre áreas mais frequentes e de que forma este padrão varia com o tempo de permanência em loja ou com o número de zonas da loja visitadas.

4 Descrição da solução implementada

Para dar resposta à necessidade apresentada no capítulo 2, foram desenvolvidos dois modelos de análise, ambos em *Excel VBA*. O primeiro diz respeito à obtenção dos trajetos efetuados pelos clientes, a partir dos valores de potência do sinal registados pelas antenas instaladas, e o segundo permite uma análise detalhada desses mesmos trajetos.

Desta forma, mesmo que a empresa, no futuro, opte por mudar a forma como obtém a posição dos clientes em loja, poderá continuar a utilizar o modelo de análise de trajetos na etapa seguinte. Para isso, apenas terá que garantir que o formato dos dados de entrada corresponde ao implementado.

4.1 Modelo de Posicionamento

O modelo de posicionamento foi dividido em cinco fases. O objetivo foi automatizar o processo e garantir que informação relativa a qualquer tipo de loja poderia ser transformada em trajetos válidos.

Neste modelo, os dados iniciais são relativos às leituras da potência de sinal de cada endereço MAC, em cada momento, e o resultado final corresponde às posições ocupadas por cada cliente, tendo em conta restrições definidas ao longo do processo.

O facto de se estar a processar uma elevada quantidade de dados fez com que houvesse a necessidade de se dar especial atenção ao tempo de processamento dos mesmos.

Inicialmente, os dados resultantes de cada etapa eram mantidos no modelo de posicionamento, aumentando, assim, o tamanho desse ficheiro. Á medida que novas folhas iam sendo criadas, a memória disponível diminuía e o tempo de processamento aumentava. A solução passou por guardar os resultados de cada etapa em ficheiros separados, numa localização especificada pelo utilizador.

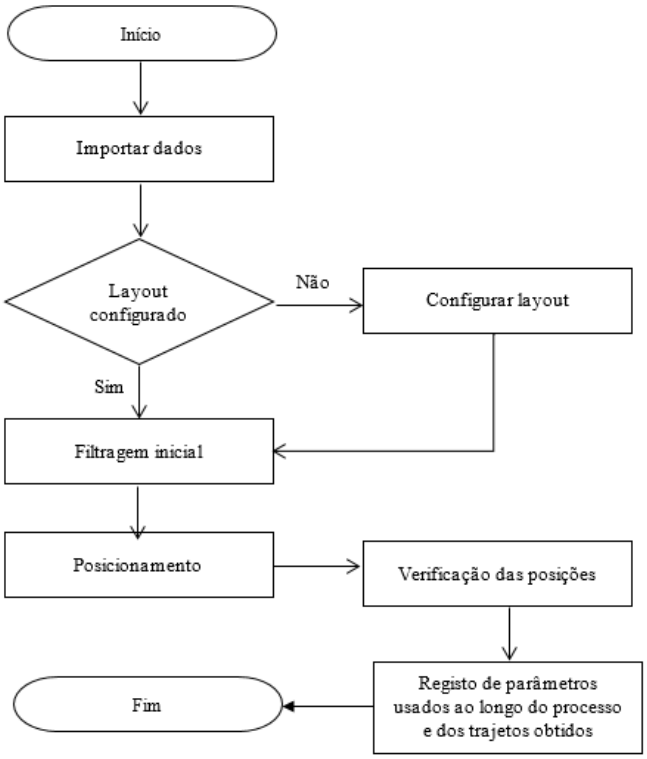


Figura 4 - Fluxograma do processo de posicionamento.

Em cada etapa, decide-se se se pretende que os passos seguintes sejam também executados. No início de cada projeto poderá ser necessário avaliar os resultados de cada fase e ajustar os parâmetros inseridos. É por isso que é dada a possibilidade de se executar cada etapa de forma independente.

Cada etapa presente na figura 4 é descrita de forma detalhada nas secções subsequentes.

4.1.1 Acesso à base de dados e seleção da informação

O objetivo desta primeira fase é obter a informação a ser analisada diretamente da base de dados da InovRetail. Na obtenção dos dados de forma automática, é necessário definir as credenciais que permitem o acesso ao *SQL Server* (*host*, *user*, *password* e *database*) e ainda os critérios que permitem a seleção dos dados a serem estudados, tais como o id associado à loja, o intervalo de tempo pretendido (dia e horas) e o tipo de antena (*Wi-Fi* ou *Bluetooth*). O utilizador deverá ainda especificar qual a localização onde pretende guardar a informação obtida.

Na Figura 1, é possível observar parte do modelo de posicionamento que diz respeito a esta etapa. A caixa de seleção "All Steps" está relacionada com o processamento de todas as etapas. Caso não esteja ativa, apenas este passo é executado quando se clica no botão "RUN".

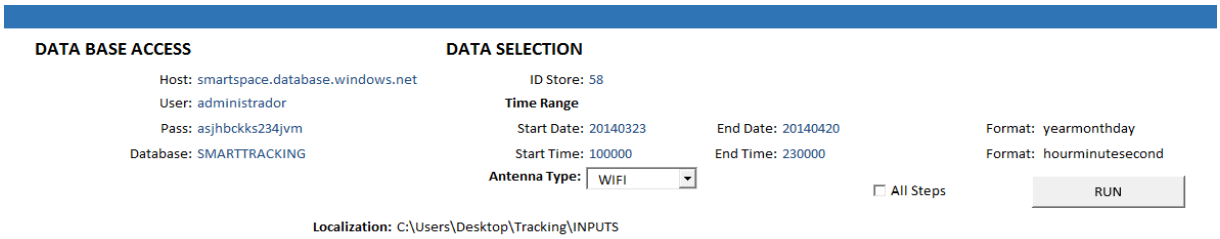


Figura 5 – *Screenshot* da parte do modelo de posicionamento relativa ao acesso à base de dados e à especificação da informação a ser extraída.

Os campos extraídos dizem respeito ao id da loja, ao endereço MAC, à data e hora, ao id da antena e à potência do sinal da leitura. É ainda extraído um campo com nome SS_INTERVAL que corresponde ao intervalo de 5 segundos desse registo.

Tal como foi previamente identificado, um dos problemas desta etapa é o facto de o número total de registos a ser estudado ultrapassar, frequentemente, o limite de linhas por folha de cálculo do *Excel*. Para superar este desafio, começou-se por decidir que, independentemente do período de tempo escolhido, iria ser criada uma nova folha de cálculo por cada dia.

Apesar de, nos dados testados, não ter havido um único dia cujo total de registos excedesse esse limite, foi necessário garantir que caso isso acontecesse a informação seria conseguida de outra forma. Assim, em cada dia é feita uma *query* que retorna o número total de registos desse período. Caso esse total se encontre dentro do limite, os dados que cumpram as restrições introduzidas pelo utilizador são importados do *SQL Server* para uma folha de cálculo criada previamente. Se o número total de registos ultrapassar o limite máximo referido, irá ser feita uma nova consulta que retorna todos os endereços MAC diferentes registados naquele dia e a consulta será feita endereço a endereço. Sempre que o limite esteja próximo, é criada uma nova folha de cálculo. Só desta forma se garante que um mesmo endereço MAC não fica em folhas de cálculo diferentes.

Nos dois casos, os dados são posteriormente ordenados por endereço MAC e por hora. Este agrupamento é feito para facilitar as análises seguintes. A ordenação não é efetuada diretamente na *query* uma vez que demora mais tempo do que a ordenação implementada em *VBA*.

4.1.2 *Configuração de Layout*

Um dos problemas identificados anteriormente está relacionado com a necessidade de se ter em conta as restrições de *layout* da loja e/ou a posição de cada antena para se conseguir obter os movimentos dos clientes.

As características de cada espaço são variáveis e, para que a ferramenta desenvolvida possa ser aplicada transversalmente, é essencial que estas possam ser introduzidas no modelo. Para dar resposta a essa necessidade, foi criada esta etapa cujo objetivo é configurar o *layout* de uma loja.

Como este tipo de informação é necessário para as fases seguintes do processo, optou-se por criar uma nova folha de cálculo, que permanecerá no ficheiro, sempre que se pretende analisar um novo espaço. Desta forma, este passo é necessário apenas se nunca tiver sido criado uma folha com as especificações da loja em análise. Nessa folha de *layout* serão registadas as antenas existentes na loja, as suas posições e a área a que correspondem.

Com o id da loja, introduzido pelo utilizador, é feita uma consulta à base de dados que retorna o tipo de antena e o id de cada antena presente nessa loja. A posição das antenas e a área a que cada uma corresponde terá posteriormente de ser introduzida no modelo uma vez que não existe essa informação na base de dados.

É ainda criada uma matriz por cada antena registada e uma matriz adicional para representar o *layout*. Estas matrizes terão o comprimento e a largura definidas pelo utilizador. O preenchimento da matriz relativa ao *layout* da loja é obrigatório para qualquer um dos métodos de posicionamento, contudo, as matrizes relativas a cada antena deverão apenas ser preenchidas caso se pretenda obter uma posição por análise de cenário.

Na Figura 6, é possível observar parte do modelo de posicionamento que diz respeito à configuração de uma loja. Os dados de acesso à base de dados, necessários para obter o id das diferentes antenas presentes no espaço, estão já especificados na etapa anterior e são usados também nesta fase.

Figura 6 - *Screenshot* da parte do modelo de posicionamento relativa à configuração do layout.

A Figura 7 apresenta um exemplo de parte de uma folha de cálculo com as características de uma loja. As áreas terão de ser previamente definidas tendo em conta o *layout* do espaço e a disposição dos produtos para que depois possam ser introduzidas no modelo.

	A	B	C	D	E	F	G	H	I	J	K
1			POSICAO					LAYOUT			
2	TYPE	ANT_ID	X	Y				1	2	3	
3	BLE	50	1	8,5			9	PROV	ARM	ARM	
4	BLE	51	1	3,5			8	M16	M16-18	M18	
5	BLE	52	3	7,5			7	M16	M16-18	M18	
6	BLE	53	3	3,5			6	POS	M20	M18-19	
7	BLE	54	2	5,5			5	POS	M20	M18-19	
8	WIFI	74	1	8,5			4	M17	M17-19	M19	
9	WIFI	75	1	3,5			3	M17	M17-19	M19	
10	WIFI	76	3	7,5			2	Montra	ENTRADA	Montra	
11	WIFI	77	3	3,5			1	Fora17	ENTRADA	Fora19	
12	WIFI	78	2	5,5							
13											
14							ANTENA	50			
15								1	2	3	
16							9	-1000	-1000	-1000	

PARAMETERS | LAYOUT_58 | LAYOUT_59 | LAYOUT_2115 | LAYOUT_1029

Figura 7 - Exemplo de parte de uma folha de cálculo com as configurações de uma loja.

4.1.3 Data Cleaning

Grande parte dos endereços MAC captados não pertence a visitantes da loja. É, por isso, fundamental eliminar registos que não trazem informação relevante. Este é o objetivo desta etapa.

Apesar deste tipo de filtragem ter sido feito em todos os projetos da InovRetail, não houve qualquer padronização do processo. Devido à elevada quantidade de registos, optou-se por fazer esta seleção mesmo antes do posicionamento. Assim, o tempo de processamento das etapas seguintes será bastante menor. Só cerca de 20% dos registos resistem a esta seleção.

Os parâmetros tidos em conta, nesta fase, dizem respeito ao número mínimo e máximo de vezes que cada endereço foi registado, ao tempo mínimo e máximo registado e ao número mínimo de antenas detetadas por cliente. Esta última restrição não tinha ainda sido implementada em projetos anteriores.

Os registos de uma pessoa que passa fora da loja são facilmente eliminados pelas restrições relativas ao tempo e número de leituras mínimo. Contudo, é necessário que se eliminem também leituras de dispositivos móveis que estão algum tempo próximo da loja mas que não respeitam a entrada na loja.

Quando uma pessoa está próxima do espaço comercial, o seu dispositivo móvel é facilmente captado pelas antenas posicionadas próximo da entrada. Porém, as antenas mais distantes não o conseguem detetar. Ao garantir que antenas mais afastadas da entrada captaram o dispositivo, aumenta-se a probabilidade de estes terem estado dentro da loja. Daí o facto de se acrescentar a restrição relativa ao número mínimo de antenas que detetaram determinado endereço MAC.

Como as antenas são equipamentos que estão sujeitos a avarias, torna-se também fundamental perceber se realmente estavam a funcionar corretamente durante o período de tempo analisado. A lista do total de antenas instaladas na loja é obtida recorrendo à folha de cálculo que contém informação referente ao *layout* e ao equipamento da loja, criada previamente. Para isso, é necessário que se especifique o tempo máximo a partir do qual se considera que a antena está com problemas na deteção. Tendo em conta esse valor, irá ser feita uma análise das leituras, e será registado o id da antena, a última hora a que foi registada e a duração da sua inatividade, caso esta apresente uma inatividade superior ao especificado. É da responsabilidade do utilizador decidir se quer incluir esses dados no estudo das posições ou se elimina por completo esse período de tempo da análise.

Optou-se por tornar todos estes parâmetros dinâmicos porque são critérios subjetivos que dependerão de quem está a fazer a análise. Desta forma, e para que o processo esteja automatizado, é essencial que estas decisões possam ser tomadas projeto a projeto, sem para isso ser necessário refazer o modelo de posicionamento.

Na Figura 8, é apresentada a parte do modelo em que se deverão especificar os parâmetros desta etapa.

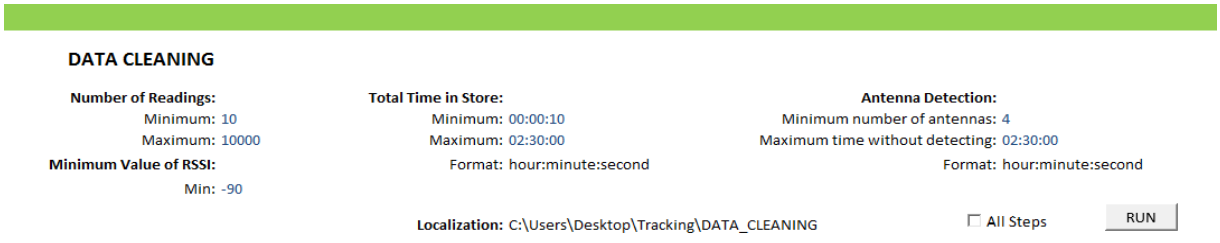


Figura 8 - *Screenshot* da parte do modelo de posicionamento relativa seleção de endereços MAC válidos.

4.1.4 Posicionamento

Depois de filtrados os dados, é altura de obter a posição dos clientes em cada momento. No desenvolvimento desta etapa, foi necessário avaliar os métodos de posicionamento aplicados anteriormente em cada projeto. Até ao momento, o trajeto de cada visitante foi obtido através de proximidade, análise de cenário e triangulação, dependendo do caso.

A triangulação foi aplicada apenas num projeto e as posições obtidas através desse método não eram mais precisas do que as obtidas através de análise de cenário, método de posicionamento usado primeiramente nesses mesmos dados. Por esta razão, foi decidido que a triangulação não seria um método implementado nesta ferramenta de análise.

Para se aplicar o método de análise de cenário, é obrigatório que se recolha informação prévia relativa ao valor de RSSI registado por cada antena, em cada posição da loja. Poderá haver casos em que não seja possível fazer o reconhecimento do local e obter esse tipo de dados. Nessas circunstâncias, será ainda possível obter o posicionamento dos clientes usando a proximidade como método de posicionamento, outro dos procedimentos implementados neste modelo.

No método de proximidade, a posição é atribuída com base na posição da antena que registou um menor valor naquele instante.

No método de análise de cenário, cada antena terá uma matriz de potências atribuída, que permite saber a potência lida numa posição com coordenadas x e y. Todas as posições são testadas e a posição atribuída àquele endereço MAC, naquele momento, será aquela que apresente um erro quadrático médio menor, no conjunto das antenas.

Tal como se pode ver na Figura 9, o utilizador deverá especificar o método aplicado no processo de posicionamento.

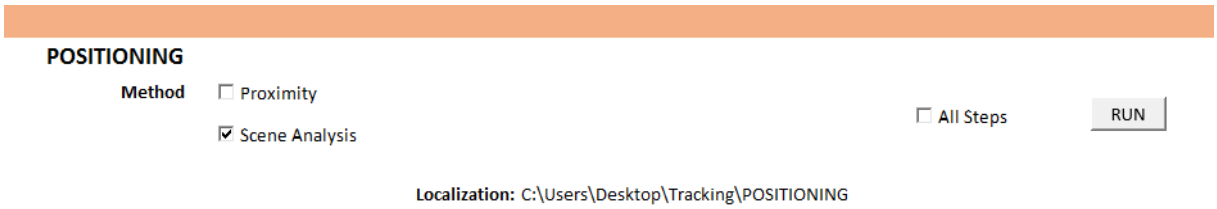


Figura 9 - *Screenshot* da parte do modelo de posicionamento relativa à descodificação dos valores de RSSI.

Independentemente do método escolhido, o primeiro passo desta etapa é a criação de uma tabela *pivot* que organiza leituras efetuadas por um mesmo endereço MAC, num determinado momento, todas na mesma linha. Na Figura 10 é possível observar a estrutura dos dados antes e depois desta etapa. Quanto menor o número de linhas a ser percorrido, menor o tempo envolvido na execução do processo. Esta é uma das vantagens de se alterar a forma como é distribuída a informação, no *Excel*.

São os passos seguintes que diferem consoante o método de posicionamento adotado.

Em relação ao posicionamento por proximidade, o processo é simples. Os valores de RSSI lidos por cada antena, em cada momento, são analisados e é registado o máximo desses valores. Tendo em conta esse valor, é depois ainda mantido em memória o id da antena que o registou. O passo seguinte passa por consultar a posição dessa antena para depois atribuir essa mesma posição àquele cliente, naquele instante. Tal como já foi referido, a localização de cada antena está descriminada na página referente às configurações do *layout* dessa loja, folha de cálculo criada previamente.

	A	B	C	D	E	F	G	H	I	J
1	ORG_DETAIL_ID	MAC_ID	ANT_ID	DATE_ID	TIME_ID	SS_INTERVAL	READ_VALUE_RSSI			
2	59	00:02:78:B0:2D:05	62	20141028	14:47:00	0	-60			
3	59	00:02:78:B0:2D:05	65	20141028	14:47:00	0	-64			
4	59	00:02:78:B0:2D:05	66	20141028	14:47:00	0	-64			
5	59	00:02:78:B0:2D:05	68	20141028	14:47:00	0	-68			
6	59	00:02:78:B0:2D:05	70	20141028	14:47:00	0	-60			
7	59	00:02:78:B0:2D:05	71	20141028	14:47:00	0	-57			
8	59	00:02:78:B0:2D:05	72	20141028	14:47:00	0	-51			
9	59	00:02:78:B0:2D:05	64	20141028	14:47:05	1	-61			
10	59	00:02:78:B0:2D:05	67	20141028	14:47:20	4	-68			
11	59	00:02:78:B0:2D:05	69	20141028	14:47:20	4	-64			

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Média de READ_VALUE_RSSI			ANT_ID												Position	
2	DATE_ID	MAC_ID	TIME_ID	62	64	65	66	67	68	69	70	71	72	Máx RSSI	Ant_ID	X	Y
3	20141028	00:02:78:B0:2D:05	14:47:00	-60		-64	-64		-68		-60	-57	-51	-51	72	5	2
4	20141028	00:02:78:B0:2D:05	14:47:05		-61									-61	64	1	1,5
5	20141028	00:02:78:B0:2D:05	14:47:20					-68		-64				-64	69	4	4

Figura 10 - Exemplo da disposição dos dados antes e depois da etapa de posicionamento por proximidade.

O processo de posicionamento por análise de cenário é mais complexo. Para ser bem sucedido, foi definido um novo tipo de dados Matriz_Antena, que corresponde ao id da antena e à sua matriz de valores de RSSI.

Quando se seleciona este método, é criado um *array* de Matriz_Antena que é preenchido à medida que se percorre as matrizes de cada antena, definidas na folha de cálculo onde se configurou o *layout* da loja.

No próximo passo, e para cada linha, são gravados os ids das antenas que registaram o endereço MAC naquele momento e o respetivo valor RSSI. De seguida, para cada posição possível de uma matriz, é calculado o erro quadrático do valor lido e do valor previamente mapeado. A posição atribuída será aquela que, no conjunto de todas as antenas que registaram o dispositivo naquele momento, apresente um menor erro quadrático médio.

4.1.5 Pós-processamento

Depois de obtidas as posições, é essencial que se faça uma avaliação dos trajetos. Um dos critérios usados até ao momento está relacionado com o facto de, no conjunto de posições atribuídas a cada endereço MAC, se verificar se há registo de entrada e saída da loja.

Além disso, é ainda investigado se a posição seguinte é possível tendo em conta a distância percorrida nesse intervalo de tempo. Outro dos parâmetros averiguados é se a transmissão de sinal do dispositivo móvel foi interrompida em algum momento do trajeto.

A restrição relativa à distância percorrida é implementada através do número de células. Como a unidade não está definida, partirá do utilizador perceber qual é o máximo de células que podem ser percorridas por segundo.

Para calcular a distância entre duas posições é usada a expressão 4.1, relativa à distância euclidiana.

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{4.1}$$

Onde:

D, é a distância entre posições
x_1, é a coordenada do eixo dos xx da posição inicial
y_1, é a coordenada do eixo dos yy da posição inicial
x_2, é a coordenada do eixo dos xx da posição final
y_2, é a coordenada do eixo dos yy da posição final

Depois de especificado esse parâmetro, todos os registos são percorridos e caso não respeitem essa condição, a posição passará a ser uma média entre a que estava atribuída e a anterior.

Quanto à interrupção do sinal, será escrito numa nova coluna "OK" caso o trajeto esteja completo ou "NOT OK" caso o sinal tenha sido interrompido por um período de tempo superior ao especificado pelo utilizador.

Na análise de todas as posições, é também verificado se cada trajeto registou uma entrada ou saída. Caso o trajeto seja válido, relativamente a esse parâmetro, na coluna seguinte é escrito, mais uma vez, "OK" ou "NOT OK".

É ainda de referir que, nesta etapa, o primeiro passo é atribuir, a cada posição, a zona correspondente. Para isso, é considerada a matriz correspondente ao *layout* da loja, previamente preenchida com as diferentes áreas da loja.

A interface relativa a esta etapa está exemplificada na Figura 11.

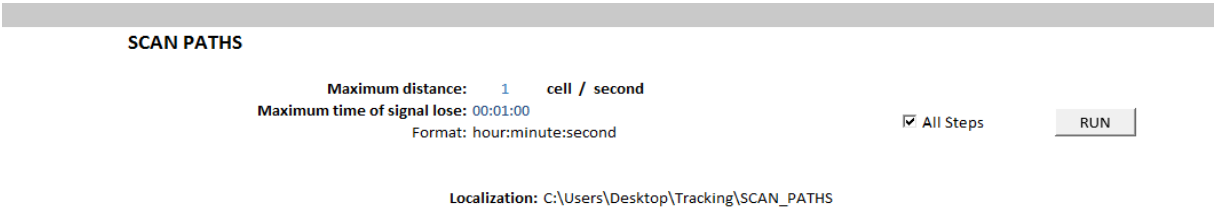


Figura 11 – *Screenshot* da etapa relativa à análise dos trajetos.

4.1.6 Output

Nesta última fase, todos os parâmetros e restrições usadas no processo são copiados para a folha de cálculo. A folha de output terá o id da loja, o tipo de antena, o método utilizado, o mínimo de leituras, o máximo de leituras, o valor de RSSI mínimo, o mínimo de tempo, o máximo de tempo, o número mínimo de antenas, o número do trajeto, a data, o ano, o mês, o dia, o dia da semana, o MAC_ID, o tempo, a hora, o minuto, o segundo, a coordenada x, a coordenada y, a zona correspondente, a percentagem de posições alteradas no trajeto, o registo de entrada/saída, e a verificação do tempo máximo de perda de sinal. O nome de cada ficheiro é composto pelo id da loja, data, e dia da semana a que corresponde.

4.2 Modelo de análise dos trajetos

Obtidos os trajetos dos clientes, foi necessário desenvolver outro modelo responsável por extrair informação relevante desses mesmos dados.

Dado que existem vários tipos de trajetos numa loja, optou-se por criar uma ferramenta de análise que permita ao utilizador decidir se pretende avaliar a loja com todos os dados disponíveis ou se pretende agrupar primeiramente os trajetos em *clusters* e só depois fazer uma análise de cada grupo de trajetos.

Neste processo, serão considerados apenas trajetos completos, isto é, trajetos em que tenha sido registado uma entrada e uma saída da loja. Além disso, não serão analisados endereços MAC que tenham um intervalo de tempo entre leituras superior ao número definido pelo utilizador no modelo de posicionamento.

Esta ferramenta permite calcular, de forma automatizada, as métricas já anteriormente estudadas pela empresa. Para além disso, será ainda possível avaliar as transições entre as diferentes áreas de cada loja, utilizando cadeias de Markov.

A Figura 12 apresenta um fluxograma do modelo de análise dos trajetos.

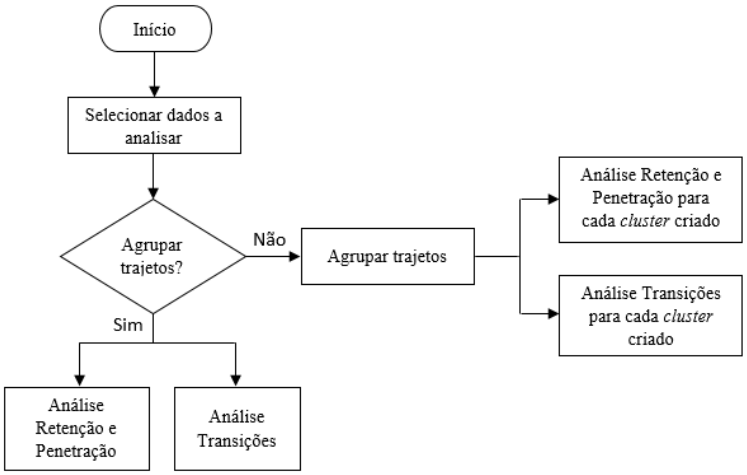


Figura 12 - Fluxograma do processo relativo à análise dos trajetos.

4.2.1 Seleção dos dados

Tal como no modelo de posicionamento, é necessário que primeiramente se defina que dados se pretendem estudar. O id da loja, o tipo de antena e o intervalo de tempo deverão ser especificados pelo utilizador. Para além disso, existe ainda a possibilidade de se escolher os trajetos com base nos dias da semana. Esta funcionalidade permite que os trajetos de apenas determinados dias da semana sejam estudados. Por exemplo, se for definido um intervalo de tempo de um mês e o dia da semana selecionado for apenas segunda-feira, irá ser feita uma análise apenas de trajetos registados nas segundas-feiras daquele mês.

Na Figura 13 é possível observar uma imagem da interface desenvolvida.

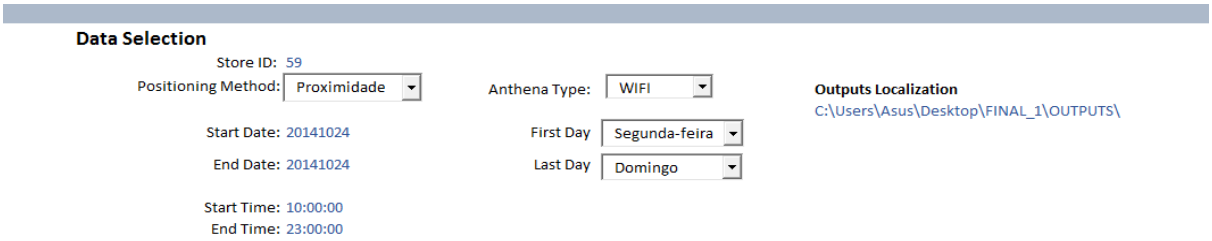


Figura 13 - Interface de análise dos trajetos: seleção de dados.

4.2.2 Análise global dos trajetos

Através deste modelo, poderá ser feito um estudo descritivo de cada trajeto válido. Este estudo engloba o tempo total de permanência em loja, o tempo de permanência em cada uma das áreas da loja, o número de áreas visitadas e o número de transições de cada trajeto. É ainda verificado se o cliente passou mais de 50% do tempo da sua visita numa determinada área (esta percentagem poderá ser alterada pelo utilizador).

Depois de obtida esta informação, o utilizador poderá usar os filtros que pretender para retirar conhecimento da mesma. É possível a partir daqui calcular o número de pessoas que visitaram a loja, em determinado período, durante menos de cinco minutos ou ainda o número de pessoas que estiveram numa zona específica, por exemplo.

A Figura 14 apresenta um exemplo dos resultados obtidos com este tipo de análise. Para cada id do trajeto é apresentado o tempo de permanência em cada uma das zonas da loja (UN51,

UN52, UN53, UN54 e UN55) e o tempo total em loja, contabilizado a partir do momento em que é registada uma entrada até ao momento em que se dá a saída do espaço. De seguida, apresenta-se o número de zonas visitadas pelo cliente e o número de transições efetuadas entre as mesmas. Tendo em conta a soma do tempo de permanência em cada uma das zonas, é verificado se o cliente permaneceu mais de 50% (ou outro valor especificado pelo utilizador) numa dessas zonas. Caso a resposta seja afirmativa, será registado um 10, caso contrário, esse valor será 0.

	A	B	C	D	E	F	G	H	I	J
1	DATE_MACID_ROUTE	UN51	UN52	UN53	UN54	UN55	TOTAL	NumAREAS	NUM TRANSICOES	>50% numa zona
2	20141024_00:07:88:D5:FE:E7_11	00:15:15	00:00:15	00:09:55	00:07:25	00:03:00	00:46:00	5	53	0
3	20141024_00:0C:E7:96:22:FB_43	00:00:30	0	00:00:00	00:00:10	0	00:03:05	2	8	10
4	20141024_00:66:4B:1F:C1:26_59	00:04:10	00:00:00	00:00:00	00:00:40	00:00:15	00:08:25	3	30	10
5	20141024_00:66:4B:46:38:90_60	00:00:55	0	00:00:00	00:00:30	00:00:05	00:16:55	3	37	0
6	20141024_00:AA:70:85:16:67_66	00:00:10	00:00:05	00:00:10	00:00:10	00:00:00	00:05:00	4	27	0
7	20141024_00:B3:09:F1:20:AC_67	00:00:00	0	0	00:02:10	0	00:02:50	1	3	10
8	20141024_04:46:65:E0:1C:90_86	00:00:00	00:00:00	00:00:00	00:00:05	00:00:00	00:01:25	1	4	10

Figura 14 - Análise global dos trajetos.

4.2.3 *Clustering dos trajetos*

O objetivo desta etapa é agrupar trajetos semelhantes com auxílio do *software RapidMiner*.

Depois de todos os trajetos terem sido descritos com a funcionalidade descrita no ponto 4.2.2, é altura de usar esses dados no processo de *clustering*. Os parâmetros a serem tidos em conta serão o tempo total da visita, o número de áreas que visitou e o facto de ser um trajeto igualmente distribuído por todas as zonas ou focado mais numa zona específica. Este último parâmetro está relacionado com a variável previamente calculada, apresentada na Figura 14, "<50% numa zona".

O algoritmo de *clustering* usado será o *k-means* e o número de clusters deverá ser previamente definido pelo utilizador. Os resultados deste agrupamento deverão ser guardados num ficheiro *Excel* para que possam, depois, ser usados no modelo de análise desenvolvido.

Na Figura 15 é possível observar o esquema do processo implementado no *RapidMiner*.

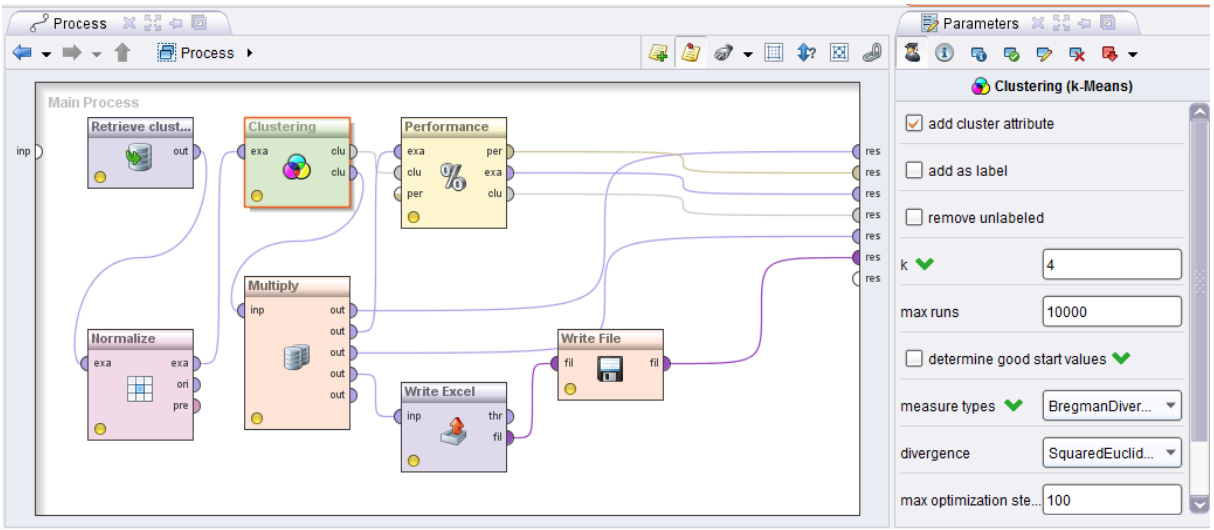


Figura 15- Esquema do processo no RapidMiner.

4.2.4 *Penetração e Retenção*

Uma das funcionalidades deste modelo é o cálculo da penetração e retenção da loja. Neste procedimento, tal como no modelo de posicionamento, é essencial haver uma folha de cálculo com as características das antenas e do *layout* do espaço.

No início do processo, é criado um vetor com as diferentes áreas da superfície comercial (com base na folha de cálculo que diz respeito ao layout). De seguida, os ficheiros correspondentes aos dias selecionados vão sendo percorridos e apenas contabilizados registos que pertençam ao intervalo de hora especificado previamente. Essa restrição tem a hora de registo de entrada na loja de cada endereço MAC como parâmetro de verificação.

A penetração é um indicador que permite saber quais as zonas da loja mais e menos acedidas. Para cada área, e em cada trajeto válido, é contabilizado se o cliente esteve naquela posição. Este número é depois dividido pelo número total de trajetos válidos e é assim obtida a proporção de pessoas que visitou determinado espaço da loja.

Na retenção, o que é medido é o tempo médio que os consumidores estão em determinada área. Este indicador é calculado somando o tempo total que os consumidores estiveram em cada zona e dividindo esse valor pelo número de trajetos válidos analisados.

Os resultados são apresentados numa nova folha de cálculo. Um exemplo poderá ser observado na Figura 16.

	A	B	C	D	E	F	G	H
1	PERIODO DE ANÁLISE				ZONE	PENETRAÇÃO	RETENÇÃO	
2	Data inicio	20141024			UN51	20,74%	00:01:36	
3	Data fim	20141031			UN52	9,20%	00:07:40	
4	Primeiro dia	Segunda-feira			UN53	23,41%	00:03:04	
5	Ultimo dia	Domingo			UN54	27,23%	00:00:26	
6	Hora Inicio	10:00:00			UN55	19,41%	00:02:14	
7	Hora Final	23:00:00						
8								
9	TOTAL VISITANTES	303						
10	MAC_IDS REGISTADOS	7535						
11								
12								

Figura 16 - Resultados de Penetração e Retenção.

Tal como já foi referido, é possível fazer uma análise destas métricas por *cluster*. Caso seja esta a opção, será criada uma folha de cálculo com os resultados para cada grupo de trajetos.

4.2.5 Análise de transições

Depois de agrupados os trajetos, é importante perceber de que forma se diferenciam. Para esse fim, são utilizadas cadeias de Markov.

Cada *cluster* irá ser analisado separadamente e segundo dois cenários diferentes. Numa das situações, a entrada e a saída são tidas em conta e, na outra, são apenas consideradas áreas referentes ao interior da loja. Estudando as transições em cada um dos casos, facilmente se percebe que no primeiro obter-se-á uma cadeia de Markov absorvente e no segundo uma cadeia de Markov regular.

Com base na cadeia de Markov absorvente, é possível não só verificar a probabilidade de transição de uma área para outra como também é calculado o número de transições médias realizadas até à saída.

Na cadeia de Markov regular obtida é possível calcular a percentagem de ocupação de cada zona através do seu vetor de equilíbrio.

Nos dois casos, serão apenas consideradas transições entre áreas adjacentes da loja no cálculo da probabilidade de transição.

Ambas as funcionalidades foram implementadas em *VBA*, permitindo assim que essa análise seja feita diretamente no modelo de análise de trajetos.

Ao fazer este tipo de estudo para cada *cluster*, serão evidenciadas as diferenças existentes em cada tipo de trajeto e poder-se-á perceber melhor os padrões de deslocamento dos clientes.

Na Figura 17 é possível observar um exemplo de uma matriz de transições obtida através do modelo desenvolvido. É obtido também o vetor de equilíbrio, calculado caso a matriz seja regular.

	A	B	C	D	E	F	G
1		UN51	UN52	UN53	UN54	UN55	ENTRADA
2	UN51	0,00	0,00	0,40	0,00	0,17	0,43
3	UN52	0,00	0,00	0,85	0,15	0,00	0,00
4	UN53	0,61	0,13	0,00	0,11	0,15	0,00
5	UN54	0,00	0,25	0,10	0,00	0,30	0,35
6	UN55	0,13	0,00	0,13	0,31	0,00	0,43
7	ENTRADA	0,47	0,00	0,00	0,41	0,12	0,00
8							
9							
10							
11	Vi=	0,23	0,06	0,18	0,16	0,14	0,22
12							
13							

Figura 17 - Exemplo de uma matriz de transições e do vetor de equilíbrio obtido através do modelo implementado.

5 Descrição das amostras de teste e resultados obtidos

Neste capítulo irá ser feita uma breve descrição dos dados usados para a criação da ferramenta de análise. Depois de identificados os parâmetros usados em cada teste, serão ainda analisados os resultados obtidos.

5.1 Descrição das amostras

Para se desenvolver os modelos de análise já apresentados, foram usados dados relativos a dois espaços comerciais com características bastante diferentes. Desta forma, foi possível testar se a ferramenta criada conseguia lidar com as diferentes restrições de *layout* de cada uma das lojas.

Uma das amostras de dados diz respeito aos trajetos de clientes numa loja dedicada à comercialização de roupa, calçado e acessórios de desporto. A loja em análise está localizada num dos maiores centros comerciais de Portugal e o seu horário de funcionamento é das 10 horas da manhã até às 23 horas da noite, todos os dias da semana. Na Figura 18 é possível observar uma planta do espaço e ainda o tipo de artigos expostos em cada zona. É também identificada a entrada do espaço.

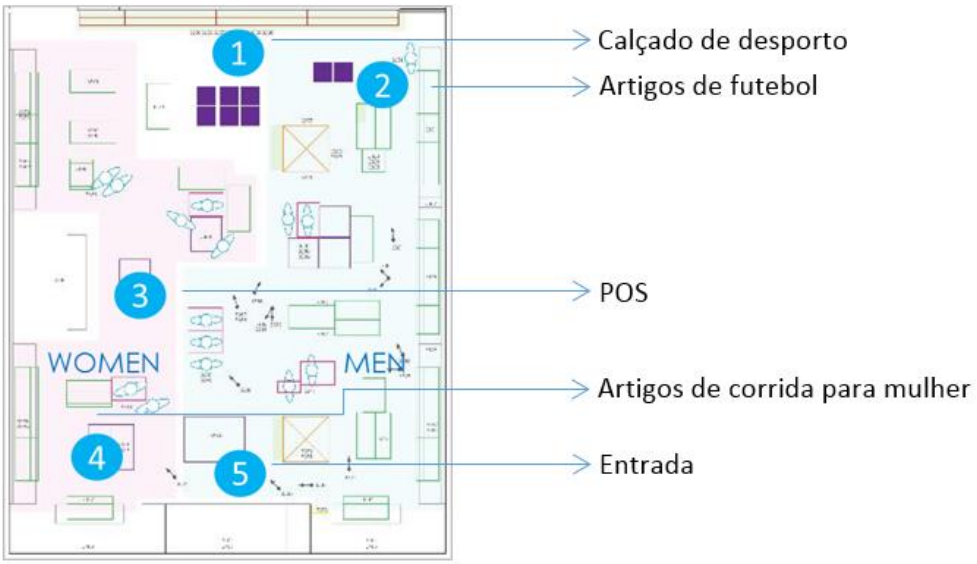


Figura 18 - *Layout* da loja de desposto analisada.

A segunda amostra de dados pertence a uma loja que se dedica à comercialização de eletrodomésticos, telecomunicações e informática e está localizada em Madrid, Espanha.

Os dados analisados dizem respeito a 8 dias e, tal como na amostra anterior, o horário de funcionamento da loja é das 10h da manhã às 23h da noite.

Em ambas as lojas, os dados relativos à potência do sinal dos dispositivos móveis dos clientes foram recolhidos via *Wi-Fi*.

A Figura 19 apresenta a planta da loja em causa, assim como as diferentes áreas que a constituem.

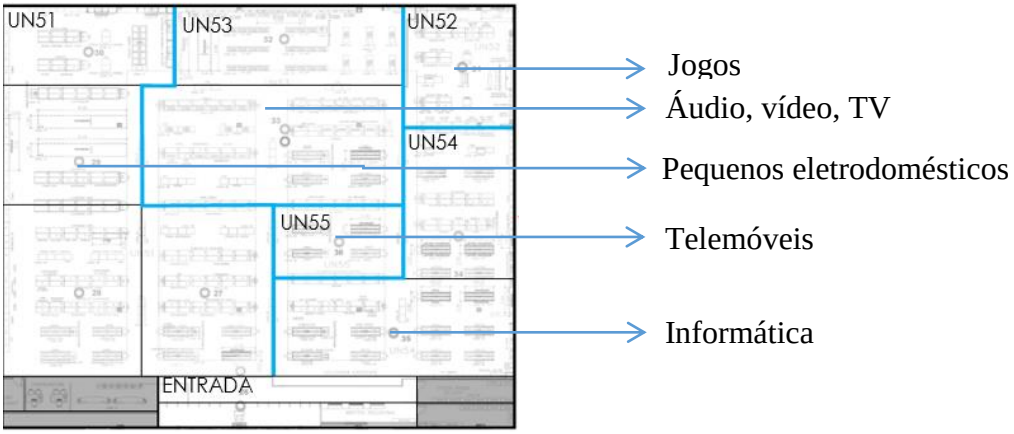


Figura 19 - *Layout* da loja de eletrodomésticos, telecomunicações e informática

5.2 Teste efetuados

Para obter os dados relativos à posição dos clientes, foi necessário definir parâmetros no *template* de posicionamento.

Em relação à filtragem inicial, foram eliminados endereços MAC cujo seu total de leituras não ultrapasse os 10 segundos e ainda os que ultrapassam as 2h e 30. Isto porque se o tempo total não supera os 10 segundos, certamente a pessoa não entrou na loja. Trajetos com leituras superiores a 2h e 30 minutos são trajetos que também não interessam estudar uma vez que podem não corresponder a clientes da loja. O trajeto comum de um cliente, numa compra, mesmo nas lojas maiores, não deverá ser superior a este tempo.

Além disso, cada endereço MAC deverá ter pelo menos 10 registos e, no máximo, 10000. Estas restrições permitem também eliminar os casos em que as pessoas não entraram na loja ou então não são clientes. É necessário ainda que cada endereço tenha sido registado em 4 antenas diferentes.

Foi necessário criar também uma folha de cálculo com informação relativa ao *layout*, às antenas e às suas posições.

Na Figura 20 e na Figura 21 é possível observar como foram registados os *layouts* da loja 1 e da loja 2, respetivamente.

G	H	I	J	K
	LAYOUT			
	1	2	3	
9	PROV	ARM	ARM	
8	M16	M16-18	M18	
7	M16	M16-18	M18	
6	POS	M20	M18-19	
5	POS	M20	M18-19	
4	M17	M17-19	M19	
3	M17	M17-19	M19	
2	Montra	ENTRADA	Montra	
1	Fora17	ENTRADA	Fora19	

Figura 20 - Representação do *layout* da loja 1.

G	H	I	J	K	L	M	N	O
	LAYOUT							
	1	2	3	4	5	6	7	8
6	UN51	UN51	UN51_1	UN53	UN53	UN53	UN52	UN52
5	UN51	UN51	UN53	UN53	UN53	UN53	UN52	UN52
4	UN51	UN51	UN53	UN53	UN53	UN53	UN54	UN54
3	UN51	UN51	UN51	UN51	UN55	UN55	UN54	UN54
2	UN51	UN51	UN51	UN51	UN55	UN55	UN54	UN54
1	UN51	UN51	ENTRADA	UN51	UN54	UN54	UN54	UN54

Figura 21 - Representação do *layout* da loja 2.

Apenas na loja 1 houve um mapeamento prévio dos valores de potência registados pelas antenas em cada posição da loja. Por essa razão, esta foi a loja escolhida para testar o método de posicionamento por análise de cenário. Quanto à loja 2, todas as posições foram obtidas por proximidade.

Desta forma, aquando da configuração do *layout*, a loja 1 foi a única que exigiu o preenchimento das matrizes de potências de cada antena. Nos dois casos, foi necessário indicar a posição das antenas e indicar as diferentes zonas da loja, tal como são apresentadas nas figuras 1 e 2.

Em relação à avaliação das posições, para ambas as lojas foi definido que a distância máxima percorrida poderia ser apenas 0,2 células/segundo. Além disso, foram desconsiderados todos os trajetos que apresentassem uma interrupção de sinal, nos seus registos, com duração superior a 1 minuto.

Na análise dos trajetos optou-se por fazer, primeiramente, uma análise geral das duas amostras em relação à penetração e retenção. Aplicou-se também uma análise Markoviana em todos os dados dos espaços comerciais e só no passo seguinte é que se repetiu o processo para cada grupo de trajetos.

Para que se conseguisse agrupar todos os trajetos com características semelhantes foram criados 5 *clusters*. Em todos foi usada a distância euclidiana como critério de agrupamento. Foram feitas 10000 iterações, no algoritmo *K-means*, e a performance foi medida através da distância entre *clusters*.

5.3 Análise dos resultados obtidos

5.3.1 Penetração e Retenção

Em qualquer uma das lojas foi possível verificar que são as zonas mais afastadas da entrada as que recebem menos visitas por parte dos consumidores.

Na loja 1, é possível verificar que é a M16 a área que menos pessoas visitam. É também facilmente percetível que é a M19 a zona por onde mais pessoas passam. Apesar da zona de POS não apresentar uma taxa de penetração muito elevada, é a área em que as pessoas passam, em média, mais tempo, quando lá passam.

Na loja 2, é a UN52, área referente aos jogos, que apresenta uma taxa de penetração de apenas 15,27%. Relativamente à taxa de retenção, é a UN54, informática, a área em que as pessoas passam, em média, mais tempo.

Nas Figura 22 e Figura 24 é possível observar a percentagem de pessoas que visita cada zona da loja 1 e 2, respetivamente. A retenção em cada uma das áreas da loja 1 e 2 é apresentada na Figura 23 e na Figura 25, respetivamente.

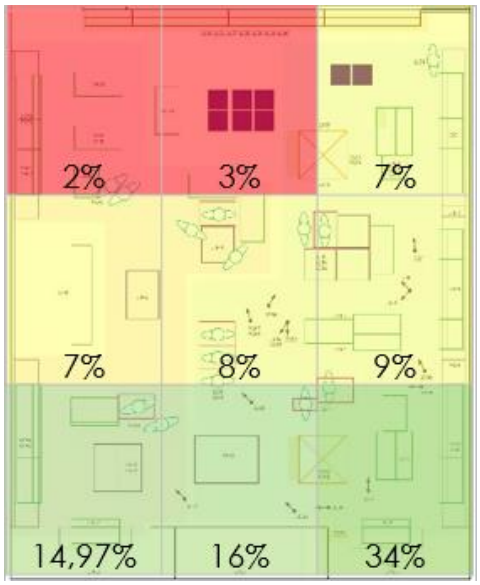


Figura 22 - Penetração da loja 1.

Figura 23 - Retenção da loja 1.

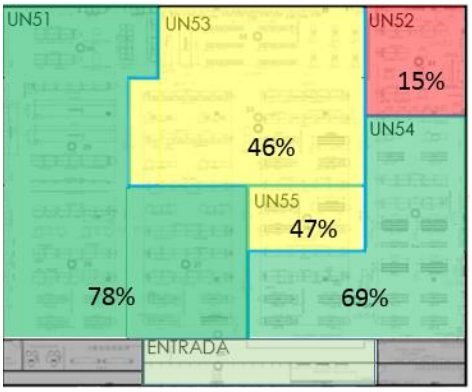


Figura 24 - Penetração da loja 2.

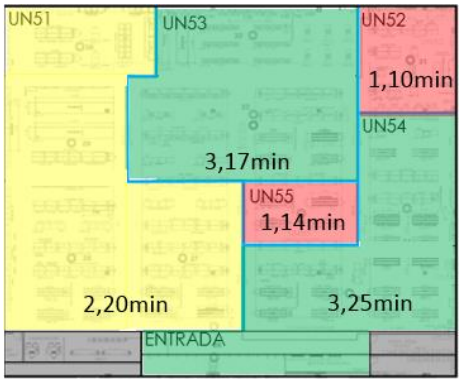


Figura 25 - Retenção da loja 2.

5.3.2 Cadeias de Markov

Para analisar as transições entre as diversas áreas, foram criados dois tipos de cadeias de Markov diferentes para cada loja, uma cadeia absorvente e outra regular. De seguida, irão ser apresentadas as matrizes de transição obtidas para cada uma das lojas bem como uma breve análise de cada um dos resultados.

Relativamente à cadeia regular, optou-se por considerar estados relativos às zonas interiores da loja, não sendo por isso considerada a entrada. Assim, a direção que as pessoas tomam à entrada de uma loja poderá ser obtida através da cadeia de Markov absorvente, onde já serão adicionados dois estados: entrada e saída.

É também apresentado um grafo das possíveis transições entre as várias áreas do espaço comercial na Figura 26. As áreas apresentadas são as previamente definidas.

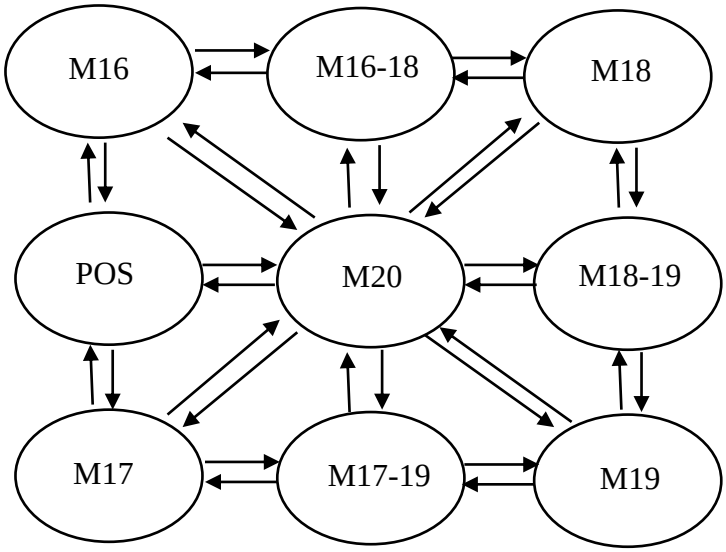


Figura 26 - Grafo representativo das transições possíveis dentro da loja 1.

Tendo em conta estas restrições, obteve-se uma cadeia de Markov com uma matriz de probabilidades de transição apresentada na Figura 27.

P_{1a} =

	M16	M17	M18	M19	M20	M16-18	M18-19	M17-19	POS
M16	0,00	0,00	0,00	0,00	**0,05**	**0,29**	0,00	0,00	**0,67**
M17	0,00	0,00	0,00	0,00	**0,21**	0,00	0,00	**0,46**	**0,33**
M18	0,00	0,00	0,00	0,00	**0,31**	**0,15**	**0,55**	0,00	0,00
M19	0,00	0,00	0,00	0,00	**0,09**	0,00	**0,42**	**0,50**	0,00
M20	**0,01**	**0,15**	**0,15**	**0,19**	0,00	**0,23**	0,00	**0,12**	**0,14**
M16-18	**0,22**	0,00	**0,17**	0,00	**0,41**	0,00	**0,09**	0,00	**0,11**
M18-19	0,00	0,00	**0,29**	**0,40**	**0,14**	**0,06**	0,00	**0,11**	0,00
M17-19	0,00	**0,24**	0,00	**0,40**	**0,13**	0,00	**0,10**	0,00	**0,13**
POS	**0,18**	**0,51**	0,00	0,00	**0,07**	**0,03**	0,00	**0,20**	0,00

Figura 27 - Matriz de transição regular da loja 1.

O vetor de equilíbrio desta matriz foi também obtido e é composto pelos valores apresentados na Figura 28.

$$V_{1i} = \begin{Bmatrix} 0,04 \\ 0,13 \\ 0,07 \\ 0,15 \\ 0,14 \\ 0,06 \\ 0,12 \\ 0,19 \\ 0,12 \end{Bmatrix}$$

Figura 28 - Vetor de equilíbrio.

Tendo em conta os resultados obtidos, percebe-se que as zonas menos visitadas, a longo prazo, são as que se encontram ao fundo da loja. Neste estabelecimento comercial é facilmente percetível que a grande parte dos trajetos se fica pelas áreas próximas da entrada. Estas são as que apresentam valores superiores no vetor de equilíbrio obtido.

Na Figura 30 está representado o esquema da loja juntamente com a probabilidade de cada área ser ocupada, no longo prazo. A verde estão apresentadas as zonas com mais probabilidade e a vermelho as que têm menos.

As probabilidades de transição que mais se distinguem das restantes foram também representadas na Figura 29.

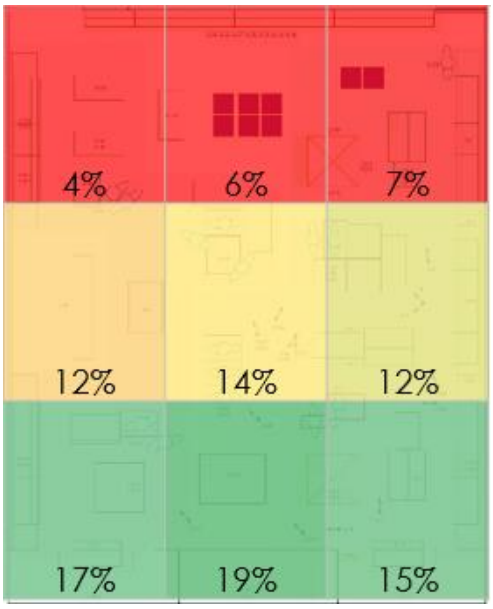


Figura 30 - Percentagem de vezes que um cliente se encontra em cada posição da loja 1, no longo prazo.

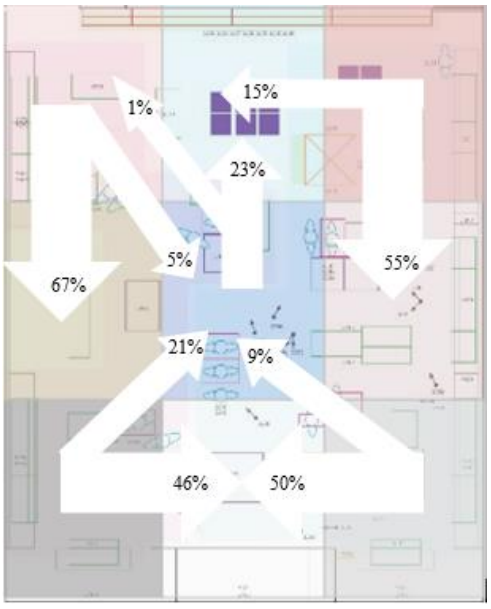


Figura 29 - Principais diferenças na probabilidade de transição entre áreas da loja 1.

Relativamente à cadeia absorvente criada, foram adicionados mais dois estados possíveis ao sistema: ENTRADA e SAÍDA. A matriz de transições obtida é apresentada na Figura 31.

P_{1b} =

	SAIDA	M16	M17	M18	M19	M20	M16-18	M18-19	M17-19	POS	ENTRADA
SAIDA	**1,00**	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
M16	0,00	0,00	0,00	0,00	0,00	**0,15**	**0,40**	0,00	0,00	**0,45**	0,00
M17	**0,26**	0,00	0,00	0,00	0,00	**0,14**	0,00	0,00	**0,40**	**0,20**	0,00
M18	0,00	0,00	0,00	0,00	0,00	**0,37**	**0,11**	**0,52**	0,00	0,00	0,00
M19	**0,55**	0,00	0,00	0,00	0,00	**0,06**	0,00	**0,15**	**0,24**	0,00	0,00
M20	0,00	**0,02**	**0,19**	**0,25**	**0,21**	0,00	**0,08**	0,00	**0,11**	**0,13**	0,00
M16-18	0,00	**0,21**	0,00	**0,29**	0,00	**0,25**	0,00	**0,11**	0,00	**0,15**	0,00
M18-19	0,00	0,00	0,00	**0,31**	**0,44**	**0,12**	**0,04**	0,00	**0,08**	0,00	0,00
M17-19	**0,22**	0,00	**0,25**	0,00	**0,38**	**0,06**	0,00	**0,05**	0,00	**0,05**	0,00
POS	0,00	**0,12**	**0,44**	0,00	0,00	**0,24**	**0,04**	0,00	**0,17**	0,00	0,00
ENTRADA	0,00	0,00	**0,24**	0,00	**0,62**	0,00	0,00	0,00	**0,15**	0,00	0,00

Figura 31 - Matriz de transição absorvente da loja 1.

A matriz fundamental obtida está apresentada na Figura 32.

$N_1 =$

	M16	M17	M18	M19	M20	M16-18	M18-19	M17-19	POS	ENTRADA
M16	1,32	0,98	0,76	0,95	1,27	0,79	0,68	1,02	1,12	0,00
M17	**0,11**	**1,61**	**0,32**	**0,67**	**0,67**	**0,17**	**0,34**	**1,00**	**0,53**	**0,00**
M18	0,17	0,68	1,84	1,15	1,30	0,44	1,22	0,88	0,49	0,00
M19	**0,05**	**0,28**	**0,24**	**1,46**	**0,36**	**0,10**	**0,39**	**0,57**	**0,16**	**0,00**
M20	0,18	0,85	0,78	1,04	1,86	0,35	0,65	0,96	0,58	0,00
M16-18	0,42	0,84	1,05	1,04	1,35	1,46	0,90	0,95	0,79	0,00
M18-19	0,12	0,53	0,84	1,24	0,89	0,30	1,70	0,82	0,36	0,00
M17-19	**0,07**	**0,63**	**0,28**	**0,88**	**0,50**	**0,13**	**0,38**	**1,61**	**0,32**	**0,00**
POS	0,27	1,16	0,50	0,84	1,02	0,32	0,48	1,09	1,58	0,00
ENTRADA	0,06	0,65	0,27	1,19	0,45	0,12	0,37	0,82	0,27	1,00

Figura 32 - Matriz fundamental da loja 1.

A partir da matriz fundamental é possível verificar que existem diferenças significativas no número médio de transições dependendo da direção de entrada. Uma pessoa que comece a sua visita à loja por a área M17 faz, em média, 6 transições entre as diferentes zonas da loja. Por outro lado, se a área inicial escolhida for a M19, são feitas apenas 4 transições. Quando os clientes optam por seguir em frente, partindo da M17-19, fazem-se, em média 5 transições.

Relativamente à direção que os clientes tomam à entrada, na matriz P1b pode-se verificar que mais de metade opta por virar à direita. Na Figura 33 apresenta-se uma imagem que caracteriza as direções de entrada e de saída da loja. Para calcular

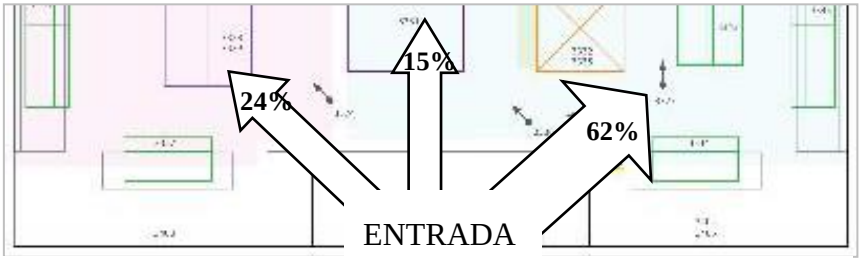


Figura 33 - Direções de entrada na loja 1.

Quanto à loja 2, é apresentado na Figura 34 um grafo com as respetivas restrições de movimento entre as várias zonas da loja.

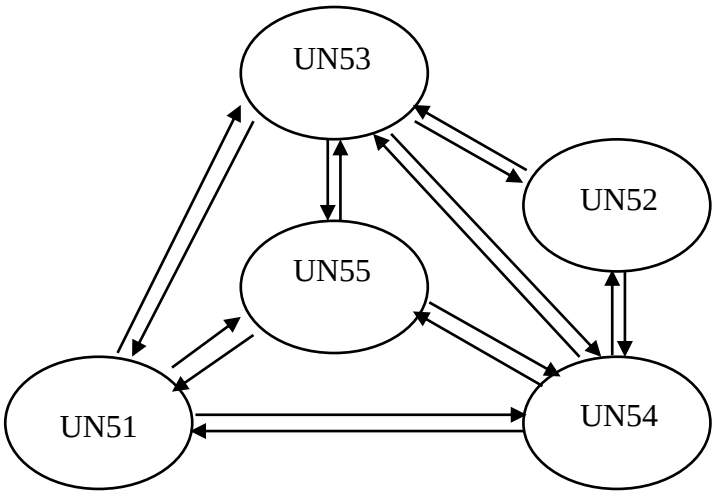


Figura 34 - Grafo representativo das transições possíveis dentro da loja 2.

A cadeia de Markov regular obtida para esta loja está traduzida na matriz de transições apresentada na Figura 35.

$P_{2a} =$

	UN51	UN52	UN53	UN54	UN55
UN51	0,00	0,00	**0,61**	**0,26**	**0,13**
UN52	0,00	0,00	**0,62**	**0,38**	0,00
UN53	**0,44**	**0,12**	0,00	**0,14**	**0,30**
UN54	0,00	**0,20**	**0,11**	0,00	**0,69**
UN55	**0,10**	0,00	**0,45**	**0,45**	0,00

Figura 35 - Matriz de transição regular da loja 2.

A Figura 36 apresenta o vetor de equilíbrio gerado a partir de P2a.

$$V_{2i} = \begin{Bmatrix} 0,15 \\ 0,08 \\ 0,28 \\ 0,23 \\ 0,26 \end{Bmatrix}$$

Figura 36 - Vetor de equilíbrio.

Neste espaço verifica-se que a UN53, onde estão expostos equipamentos de áudio, vídeo e TV, é a área onde mais pessoas estão a longo prazo, apesar de não ser a mais próxima da entrada.
A área dos jogos, UN52, continua a ser uma área com pouca percentagem de ocupação e a UN51 é também uma área mais de passagem do que de permanência.

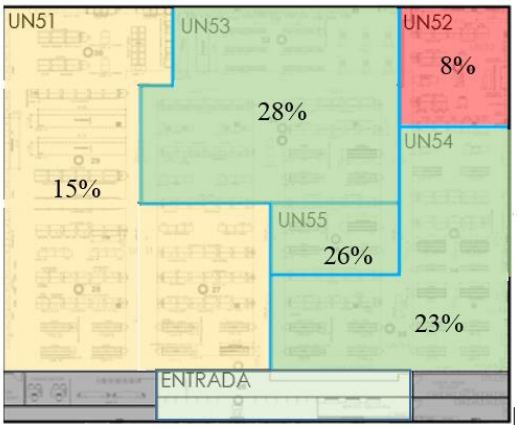


Figura 38 - Percentagem de vezes que um cliente se encontra em cada posição da loja 2, no longo prazo.

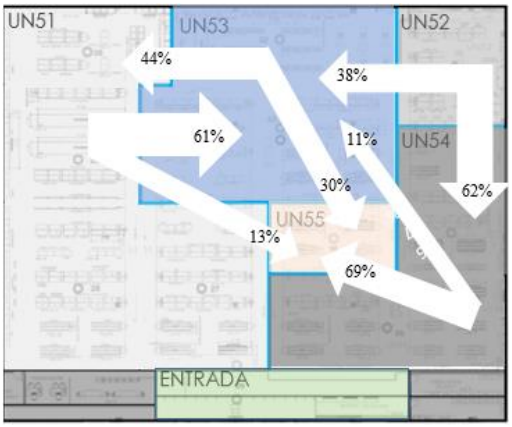


Figura 37 - Principais diferenças na probabilidade de transição entre áreas da loja 2.

Se a zona de entrada/saída da loja for considerada, a cadeia de Markov resultante é absorvente. Na Figura 39 é apresentada a matriz de transição do caso descrito.

P_{2b} =

	SAIDA	ENTRADA	UN51	UN52	UN53	UN54	UN55
SAIDA	**1,00**	0,00	0,00	0,00	0,00	0,00	0,00
ENTRADA	0,00	0,00	**0,52**	0,00	0,00	**0,48**	0,00
UN51	**0,09**	0,00	0,00	0,00	**0,56**	**0,24**	**0,12**
UN52	0,00	0,00	0,00	0,00	**0,62**	**0,38**	0,00
UN53	0,00	0,00	**0,44**	**0,12**	0,00	**0,14**	**0,30**
UN54	**0,07**	**0,35**	0,00	**0,12**	**0,06**	0,00	**0,40**
UN55	0,00	0,00	**0,10**	0,00	**0,45**	**0,45**	0,00

Figura 39 - Matriz de transição absorvente da loja 2.

Tendo em conta a matriz P_{2b} é ainda possível obter a sua matriz fundamental (N). Neste caso, os valores obtidos estão apresentados na Figura 40.

N_2 =

	ENTRADA	UN51	UN52	UN53	UN54	UN55
ENTRADA	4	6	2	7	7	6
UN51	**2**	**6**	**2**	**7**	**7**	**6**
UN52	3	5	3	8	7	6
UN53	3	6	2	9	7	6
UN54	**3**	**5**	**2**	**7**	**8**	**6**
UN55	3	5	2	8	7	7

Figura 40 - Matriz fundamental da loja 2.

A partir da matriz fundamental N_2 é possível inferir o número médio de transições até que o sistema seja absorvido, isto é, o número médio de vezes que uma pessoa transita de área para área na loja até a abandonar.

Se, por acaso, os trajetos se iniciarem na área UN51, o número médio de transições será 30. Caso se iniciem na área UN54, o número médio de transições será 31. Estes números foram calculados somando a linha referente a cada área supracitada.

Daqui pode-se concluir que o número médio de transições não é influenciado pela direção tomada à entrada.

Relativamente à direção que as pessoas seguem na entrada, verifica-se que não existe uma diferença significativa entre as percentagens obtidas na análise dos trajetos registados.

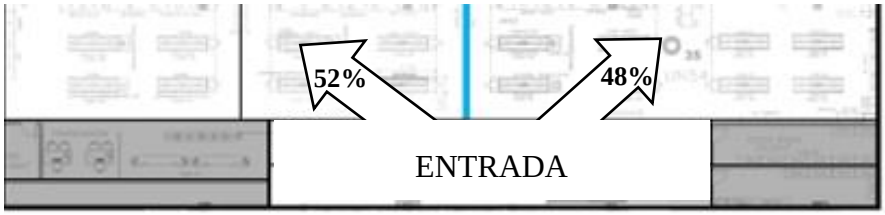


Figura 41- Direções de entrada na loja 2.

5.3.3 Clustering dos trajetos

Nesta fase serão apresentados os resultados da análise de cada grupo de trajetos obtido. Em ambas as lojas foram 5 os *clusters* criados.

Tanto numa loja como noutra, o tempo de duração do trajeto e o facto de ser ou não distribuído por várias áreas ou focado apenas numa foi o que determinou as semelhanças de trajetos.

Em relação à loja 1, a grande maioria dos trajetos tinha uma duração inferior a 6 minutos (81%). Só aproximadamente 2% das pessoas é que esteve mais de 25 minutos na loja. Apesar deste tipo de trajetos ser pouco frequente, é importante conhecer as suas características e, por isso, não foram considerados *outliers*. Além disso, 14% dos clientes estiveram até 20 minutos na loja sendo que passaram mais de metade do seu tempo apenas numa zona.

Sendo que a principal diferença entre os trajetos dos diferentes *clusters* reside no tempo de duração da visita à loja e no facto de ter estado mais do que 50% numa zona, apresentam-se na Tabela 3, as principais características dos trajetos de cada grupo.

Tabela 3 - Descrição dos trajetos de cada *cluster* la loja 1.

	Duração Visita	**Focado numa zona?**	**Proporção dos Trajetos**
Cluster 0	< 20 minutos	Sim	13,27%
Cluster 1	25 a 50 minutos	Sim e Não	0,98%
Cluster 2	< 6 minutos	Não	81,07%
Cluster 3	7 a 16 minutos	Não	4,42%
Cluster 4	> 51 minutos	Sim e Não	0,45%

Trajetos mais longos passam mais tempo na M16, M17, M16-18 e POS comparativamente aos restantes trajetos. Entende-se por trajetos longos, trajetos com uma duração superior a uma hora. Estes trajetos correspondem apenas a 0,45% dos analisados.

Trajetos curtos, com uma duração inferior a 6 minutos, que passam pelas zonas de uma forma distribuída, passam mais tempo na M19, M17-19, M18-19 e ocupam uma percentagem muito reduzida das áreas que se encontram ao fundo da loja. Estes trajetos correspondem a 81% dos trajetos analisados.

Na Tabela 4,são apresentados os vetores de equilíbrio obtidos para cada grupo de trajetos da loja 1.

Tabela 4 - Percentagem de ocupação de cada área da loja 1, a longo prazo, em diferentes tipos de trajetos.

	M16	**M17**	**M18**	**M19**	**M20**	**M16-18**	**M18-19**	**M17-19**	**POS**
Cluster 0	1,68%	13%	9%	18%	14%	3%	**12%**	24%	7%
Cluster 1	3,60%	**6%**	**28%**	**4%**	**23%**	7%	8%	12%	9%
Cluster 2	**1,08%**	11%	11%	**21%**	13%	**2%**	**12%**	**24%**	**4%**
Cluster 3	3,37%	10%	21%	12%	22%	6%	10%	10%	6%
Cluster 4	**7,36%**	**21%**	**2%**	**4%**	**12%**	**12%**	**4%**	**9%**	**29%**

Relativamente à loja 2, verifica-se que quase metade dos trajetos recolhidos corresponde a um tempo de permanência na loja inferior a 10 minutos. Além disso, apenas 3% correspondem a um tempo total de visita superior a 30 minutos.

Na Tabela 5, são apresentadas as características dos trajetos pertencentes a cada *cluster*.

Tabela 5 - Descrição dos trajetos de cada *cluster* la loja 2.

	Duração Visita	**Focado numa zona?**	**Proporção dos Trajetos**
Cluster 0	< 26 minutos	Não	37,99%
Cluster 1	31 a 90 minutos	Sim e Não	3,23%
Cluster 2	> 90 minutos	Sim e Não	1,08%
Cluster 3	<10 minutos	Sim	45,16%
Cluster 4	10 a 30 minutos	Sim	13,62%

É possível verificar que em trajetos muito longos, as pessoas acabam por estar mais tempo na UN52, uma zona com pouca percentagem de ocupação em trajetos de menor duração. Neste grupo de trajetos, é possível perceber também que a UN51 e UN53 passam a ser zonas quase só de passagem.

É também possível inferir dos resultados que em trajetos de média duração, 20 a 30 minutos, a UN53 é a que apresenta uma maior taxa de ocupação. Ao contrário do que acontece em trajetos com duração inferior a 10 minutos, em que a área mais ocupada é a UN54.

Na Tabela 6, são apresentados os vetores de equilíbrio obtidos para cada grupo de trajetos da loja 2.

Tabela 6 - Percentagem de ocupação de cada área da loja 2, a longo prazo, em diferentes tipos de trajetos.

	UN51	**UN52**	**UN53**	**UN54**	**UN55**
Cluster 0	13%	9%	18%	14%	3%
Cluster 1	**6%**	**28%**	4%	**23%**	7%
Cluster 2	11%	11%	**21%**	13%	**2%**
Cluster 3	10%	21%	12%	22%	6%
Cluster 4	**21%**	**2%**	**4%**	**12%**	**12%**

5.3.4 Sugestão de melhorias

Tendo em conta a informação obtida, existem algumas medidas que podem ser implementadas de modo a alterar a forma como os clientes se deslocam na loja.

Na loja 1, são as zonas referentes ao calçado de desporto, M16 e M16-18, as que menos pessoas visitam. É necessário, por isso, definir estratégias que possam levar o cliente a visitar essas áreas. Uma das ações que poderá ser tomada, passa por publicitar promoções referentes aos artigos vendidos no fundo da loja, à entrada. Desta forma, o consumidor sente-se mais motivado a visitar aquelas zonas. Outra das vantagens desta estratégia é o facto de o cliente ser obrigado a passar por mais áreas quando se desloca para o fundo da loja. Como o cliente vê mais artigos, a probabilidade de compra aumenta.

Ainda em relação ao *layout* da loja 1, poderia ser vantajoso alterar a zona de pagamento para uma posição mais afastada da entrada. Desta forma, o cliente é forçado a percorrer a loja até à caixa, aumentando a possibilidade de comprar mais produtos.

É fundamental que o tempo de permanência médio do cliente em loja seja também aumentado visto que mais de 80% das pessoas visitam a loja em menos de 6 minutos. Para isso, poderá ser útil criar obstáculos no caminho, como por exemplo expositores mais pequenos, de forma a obrigar o consumidor a parar para ver.

Quanto à loja 2, a zona de jogos (UN52) apresenta uma elevada taxa de ocupação em trajetos de longa duração e poderá, por isso, ser usada para publicitar produtos da UN51 e UN53, áreas não tão visitadas nesse tipo de trajetos.

Observando a figura 38, verifica-se que o corredor da UN54, UN55 e UN53 é o mais utilizado. Uma das medidas a tomar, poderá passar por criar algum tipo de promoção cruzada entre produtos dessas zonas e produtos da UN51 ou UN55. Desta forma, as pessoas tendem a visitar as restantes áreas da loja.

Estas são algumas medidas que poderão ser tomadas, contudo, é de realçar que seria necessário uma análise aprofundada do espaço para que se identificassem outras possíveis melhorias. O objetivo deste projeto foi apenas o de análise dos dados referentes ao movimento dos consumidores em loja e, portanto, a avaliação do espaço não foi realizada.

6 Conclusões e trabalhos futuros

Sendo a InovRetail uma empresa que se dedica ao estudo do comportamento do consumidor em loja, o objetivo deste projeto foi desenvolver uma ferramenta de análise capaz de transformar valores de potência, emitidos por dispositivos móveis, em posições, e extrair conhecimento útil para o negócio em causa.

Tal como havia sido delineado, a aplicação criada permite analisar os trajetos de qualquer loja, independentemente do seu *layout*. Para verificar esta premissa, todas etapas do processo foram testadas com dados relativos a duas lojas de diferentes características e em todas o resultado foi positivo.

Relativamente à nova funcionalidade implementada, a aplicação de cadeias de Markov para descrever as transições entre as diferentes zonas de cada loja permite perceber quais são realmente os caminhos mais frequentes dos clientes. Este conhecimento poderá ser utilizado, por exemplo, na colocação de publicidade nos percursos mais frequentados. Desta forma, o consumidor poderá ser levado a percorrer outras zonas da loja que de outra forma não visitaria. Além disso, o agrupamento de trajetos com semelhanças permite também perceber quais as características mais frequentes das visitas dos consumidores à loja.

Apesar de, nestes casos, apenas serem estudados os movimentos dos consumidores, a análise dos trajetos poderá ser alargada também a funcionários. Utilizar os indicadores calculados para comparar as zonas de maior retenção dos funcionários com as dos consumidores poderá ajudar os gestores de loja a melhorar o escalonamento dos seus empregados.

Com a realização deste projeto, a InovRetail pode, a partir de agora, fazer uma análise dos movimentos dos consumidores em loja de forma muito mais eficiente uma vez que deixa de ser necessário o desenvolvimento de novos modelos individuais aplicados às diferentes características de cada loja.

No futuro, poderá ser útil para a empresa o desenvolvimento de um algoritmo capaz de identificar os trajetos mais frequentes tendo em conta todas as posições de cada percurso. Neste momento, apenas é tida em consideração a posição anterior. Além disso, poderá ser interessante ter em conta o facto de um cliente ter ou não efetuado uma compra. Desta forma, poderá ser possível prever o comportamento dos consumidores, através de técnicas de mineração de dados, associando a compra a outras características dos trajetos.

Referências

Abedi, Naeim, Ashish Bhaskar e Edward Chung. 2013. "Bluetooth and Wi-Fi MAC address based crowd data collection and monitoring: benefits, challenges and enhancement".

Beder, Christian, Alan McGibney e Martin Klepal. 2011. "Predicting the expected accuracy for fingerprinting based WiFi localisation systems". Comunicação apresentada em Indoor Positioning and Indoor Navigation (IPIN), 2011 International Conference on.

Bouet, Mathieu e Aldri L Dos Santos. 2008. "RFID tags: Positioning principles and localization techniques". Comunicação apresentada em Wireless Days, 2008. WD'08. 1st IFIP.

Cabanes, Guénaël, Younès Bennani e Frédéric Dufau-Joël. 2009. "Mining Customers' Spatio-Temporal Behavior Data Using Topographic Unsupervised Learning". Comunicação apresentada em Machine Learning and Applications, 2009. ICMLA'09. International Conference on.

Cil, Ibrahim. 2012. "Consumption universes based supermarket layout through association rule mining and multidimensional scaling". *Expert Systems with Applications* no. 39 (10):8611-8625.

Delafontaine, Matthias, Mathias Versichele, Tijs Neutens e Nico Van de Weghe. 2012. "Analysing spatiotemporal sequences in Bluetooth tracking data". *Applied Geography* no. 34:659-668.

Ellersiek, Timothy, Gennady Andrienko, Natalia Andrienko, Dirk Hecker, Hendrik Stange e Marc Mueller. 2013. "Using Bluetooth to track mobility patterns: depicting its potential based on various case studies". Comunicação apresentada em Proceedings of the Fifth ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness.

Farid, Zahid, Rosdiadee Nordin e Mahamod Ismail. 2013. "Recent advances in wireless indoor localization techniques and system". *Journal of Computer Networks and Communications* no. 2013.

Farley, J.U., and Ring. 1996. "A Stochastic Model of Supermarket Traffic Flow". *Operations Research (14)*:555-567.

Fayyad, Usama, Gregory Piatetsky-Shapiro e Padhraic Smyth. 1996. "The KDD process for extracting useful knowledge from volumes of data". *Communications of the ACM* no. 39 (11):27-34.

Jain, Anil K. 2010. "Data clustering: 50 years beyond K-means". *Pattern recognition letters* no. 31 (8):651-666.

Kaemarungsi, Kamol e Prashant Krishnamurthy. 2004. "Modeling of indoor positioning systems based on location fingerprinting". Comunicação apresentada em INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies.

Khodayari, Shahrzad, Mina Maleki e Elham Hamedi. 2010. "A RSS-based fingerprinting method for positioning based on historical data". Comunicação apresentada em Performance Evaluation of Computer and Telecommunication Systems (SPECTS), 2010 International Symposium on.

Kröckel, Johannes e Freimut Bodendorf. 2011. "Visual Customer Behavior Analysis Based on Customer Movements". *THE ICITA 2011 JOURNAL OF INFORMATION TECHNOLOGY AND APPLICATIONS ISSN 1839-0048 EDITION*:24.

Levy, M. & Weitz, B. A. 2001. "Retailing management (4th ed.)". *IRWIN: McGraw-Hill*.

Lewison, D. M. 1994. "Retailing (5th ed.)". *New York, NY: Macmillan College*

Publishing Company.

Liu, Hui, Houshang Darabi, Pat Banerjee e Jing Liu. 2007. "Survey of wireless indoor positioning techniques and systems". *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* no. 37 (6):1067-1080.

Lopes, Bruno Miguel Torres. 2014. "Algoritmos de localização com informação histórica e realimentação dos utilizadores".

Mainetti, Luca, Luigi Patrono e Ilaria Sergi. 2014. "A survey on indoor positioning systems". Comunicação apresentada em Software, Telecommunications and Computer Networks (SoftCOM), 2014 22nd International Conference on.

Millonig, Alexandra e Georg Gartner. 2008. "Shadowing-Tracking-Interviewing: How to Explore Human Spatio-Temporal Behaviour Patterns". Comunicação apresentada em BMI.

Oliveira, José Luís Martins de. 2014. "Exploração de Dados de uma Solução Mobile Wallet usando as Técnicas de Data Mining".

Phua, Peilin, Bill Page e Svetlana Bogomolova. 2015. "Validating Bluetooth logging as metric for shopper behaviour studies". *Journal of Retailing and Consumer Services* no. 22:158-163.

Raju, P Salman, Dr V Rama Bai e G Krishna Chaitanya. 2014. "Data mining: Techniques for Enhancing Customer Relationship Management in Banking and Retail Industries". *International Journal of Innovative Research in Computer and Communication Engineering* no. 2 (1).

Rice, Ronald E e James E Katz. 2003. "Comparing internet and mobile phone usage: digital divides of usage, adoption, and dropouts". *Telecommunications Policy* no. 27 (8):597-623.

Sarkar, Tapan K, Zhong Ji, Kyungjung Kim, Abdellatif Medouri e Magdalena Salazar-Palma. 2003. "A survey of various propagation models for mobile communication". *Antennas and Propagation Magazine, IEEE* no. 45 (3):51-82.

Scamell-Katz, Siemon. 2012. *The Art of Shopping*. LID Pub.

Silva, Joao André, Maria João Nicolau e António Costa. 2011. "Wifi localization as a network service".

Singh, Priyanka, Neha Katiyar e Gaurav Verma. 2014. "Retail Shoppability: The Impact Of Store Atmospherics & Store Layout On Consumer Buying Patterns".

Sorensen, Herb. 2009. *Inside the mind of the shopper: the science of retailing*. Pearson Prentice Hall.

Teknomo, Kardi. 2006. "K-means clustering tutorial". *Medicine* no. 100 (4):3.

Utsch, Patrick e Thomas Liebig. 2012. "Monitoring microscopic pedestrian mobility using bluetooth". Comunicação apresentada em Intelligent Environments (IE), 2012 8th International Conference on.

Versichele, Mathias, Liesbeth De Groote, Manuel Claeys Bouuaert, Tijs Neutens, Ingrid Moerman e Nico Van de Weghe. 2014. "Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: A case study of Ghent, Belgium". *Tourism Management* no. 44:67-81.

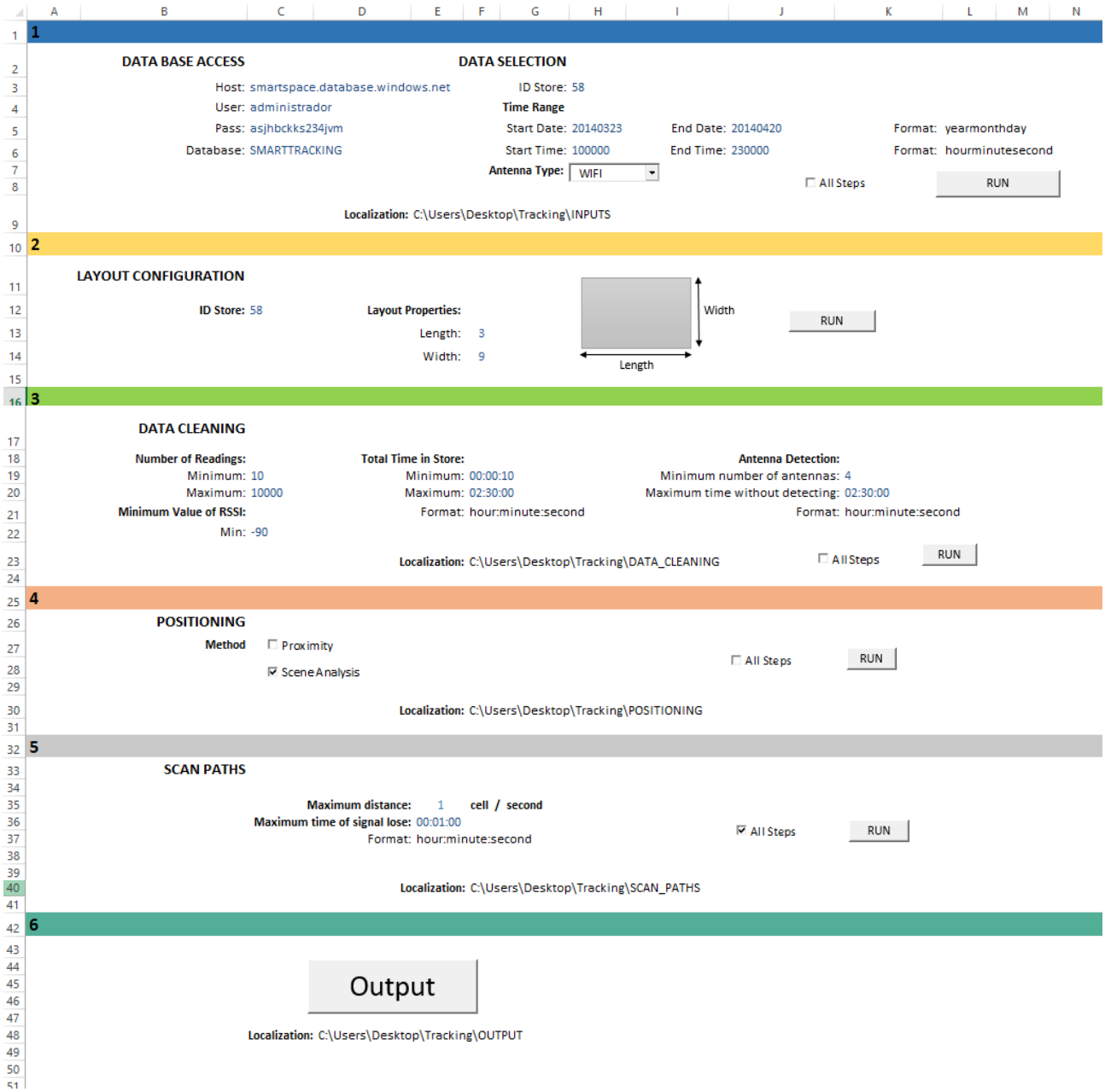
Versichele, Mathias, Tijs Neutens, Matthias Delafontaine e Nico Van de Weghe. 2012. "The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities". *Applied Geography* no. 32 (2):208-220.

Yan, Ping. 2010. "Spatial-temporal data analytics and consumer shopping behavior modeling".

Yan, Ping e Daniel D Zeng. 2008. "Clustering customer shopping trips with network structure". Comunicação apresentada em International Conference on Information Systems (ICIS).

Zhang, Da, Feng Xia, Zhuo Yang, Lin Yao e Wenhong Zhao. 2010. "Localization technologies for indoor human tracking". Comunicação apresentada em Future Information Technology (FutureTech), 2010 5th International Conference on.

ANEXO A: Interface do modelo de posicionamento



ANEXO B: Interface do modelo de análise de trajetos